Journal of Experimental Education

Volume IX DECEMBER, 1940 Number 2

THE EFFECTIVENESS OF SELECTED METHODS OF RADIO EDUCATION AT THE SCHOOL LEVEL

DONNASUE LOHMEYER AND
RALPH H. OJEMANN
Child Welfare Research Station
University of Iowa

The conviction is growing that radio may have many possibilities as a unique and valuable tool in the field of education. Educators are now asking the question, how can radio make its contributions most effectively? Are there forms or methods peculiarly suited to this means of presentation and others which are not so well adapted? In other words, what are the most effective methods of teaching by radio?

The purpose of this study was to determine the relative effectiveness of three different auditory methods for developing functional concepts and related attitudes. Three content areas, one each in history, science, and literature, were treated in three different forms and the relative effectiveness of each was measured by carefully constructed tests, using sixth grade pupils as subjects.

The selections chosen for use in this study included (1) an account of Lord Clive's Adventures in India, (2) an account of the Discovery of the Cause of African Sleeping Sickness, and (3) an adaptation of a portion of Hugo's *Les Miserables*.

The three different types of presentation to which each of these stories were adapted are: (1) straight talk, (2) informal discussion using several voices, and (3) dramatization. These were thought to represent in a general way the "first, second, and third person" methods of presentation, i. e., the actual depiction of an incident as it happens (dramatization), an eye-witness narration about the event (talk), and the third-hand discussion of it after it has been reported (discussion).

In all cases an attempt was made to equate the three scripts in each series in all respects except form. For each series a set of ten simple ideas and an attitude related to those ideas were the basis from which each of the three scripts were written.

The straight talk in each set treated the ideas in a very straight-forward, conversational, story-telling manner; the discussions in each case were built up around the situation of a school teacher conducting a very informal and unrestricted class discussion in which several students spoke at will; the dramatizations used music and sound effects freely and employed a narrator who carried the story forward from one dramatized incident to another.

When these scripts had been carefully checked in both content and technique of writing, they were cast, produced, and recorded. University students who were experienced in radio acting read the parts in the scripts, and the programs were recorded by an experienced electrical technician on 16 inch acetate discs at 33½ revolutions per minute. All programs were closely timed to 15 minutes.

The tests used to measure the effectiveness of the various programs consisted of two parts. The first was a series of multiple choice questions covering the factual ideas presented in the programs; the second part consisted of multiple-choice questions covering the corresponding attitude. For the series on Lord Clive and *Les Miserables*, these attitudes dealt with the significance of various factors determining human behavior; for the series on African Sleeping Sickness, two situations were designed for the application of the scientific attitude. Included in all of the tests was an item which rated the child's interest in each program.

In designing these tests, the basic principles of test construction were kept in mind. An attempt was made to test more than mere

"guessing power" and the words used in the construction of the tests were carefully checked with the vocabulary reported by Dale[1] as being understood by the majority of sixth grade children.

The students serving as subjects in this experiment were taken from the sixth grade classes of four schools. Groups were equated on general ability, amount of time spent listening to the radio, and the initial content score as measured by preliminary testing. Fifty subjects were used for each of the three methods.

The procedure in conducting the experiment was as follows: The tests were given to each group a week prior to the presentation of the programs; a week later the program was presented to each group in a rotating order selected by a determinant. In each case, the first program was presented, followed immediately by the test covering the material given in that program. A pause for relaxation was allowed and then the second program and the second test was given. After another relaxation period, the third program was introduced and the third test completed the testing. Following this the students were asked to give an estimate of the number of hours they had listened to the radio during the preceding week.[2]

In analyzing the data, the first question to be answered concerned the relative effectiveness of the various methods when all of the subjects were divided into three groups equated as to initial ability, use of radio, and initial content score.

In equating the groups on use of radio, the subject's estimates of the number of hours spent listening to the radio during the preceding week were used. The figures for each of the equated groups are given below:

Subject	Group	Initial Ability*	Use of Radio, Mean Hours Per Week	Initial Score**
Lord Clive	Talk	7.5	11.1	3.1
	Discussion	8.3	11.0	3.7
	Drama	7.8	11.1	3.4
African Sleeping Sickness	Talk	7.8	11.1	2.8
	Discussion	7.5	11.1	2.7
	Drama	8.1	10.6	3.1
Les Miserables	Talk	8.3	11.0	2.4
	Discussion	7.8	11.1	2.7
	Drama	7.5	11.1	2.7

*Mean grade-norms; a grade of 7.5 would mean "half-way through the seventh grade."
**Mean number answers correct out of a possible 12, 10, and 12 for Lord Clive, African Sleeping Sickness, and *Les Miserables* respectively.

The analysis of final scores was based on the gain of each subject from the score made on the initial to the final test. This was done since the purpose of the experiment was to determine which method produced the greatest change in the listener. The results of the analysis of the variance of the three groups are given below:

Subject		Degrees of Freedom	Sum of Squares	Variance	Ratio
Lord Clive	Methods	2	33.38	16.69	
	Within groups	147	971.22	6.61	2.51
	Total	149			
African Sleeping Sickness	Methods	2	38.75	19.38	
	Within groups	147	702.30	4.78	4.05
	Total	149			
Les Miserables	Methods	2	112.99	56.495	
	Within groups	147	1167.77	7.957	7.10
	Total	149			

Comparison of these ratios, for 2 and 147 degrees of freedom, with those in a table of variance ratios, shows that two of these differ-

[1] Dale, Edgar: Familiarity of 8000 common words to pupils in the fourth, sixth, and eighth grades. Bureau of Educational Research, Ohio State University. Pp. 84.

[2] Copies of all scripts and tests and the transcriptions of all the programs are found in Master of Arts theses on file in the library of the State University of Iowa.

ences are rather highly significant below the 5 per cent level. The ratio would have to be 6.81 to be significant at the 1 per cent level of confidence, and 3.91 to be significant at the 5 per cent level.

The "t-test" was applied to determine the significance of the differences between the means for the various methods within each group. The figures obtained from the application of the "t-test" are given in the following tabulation:

Difference Between Methods	Lord Clive	African Sleeping Sickness	*Les Miserables*
Talk and Discussion	1.90	1.19	3.76
Talk and drama	.19	1.65	1.63
Drama and Discussion	2.09	2.84	2.13

The "t" score for any one of these differences must equal or exceed 2.58. In the above tabulation there are two scores which exceed this value. To be significant at the 2 per cent level, the score must be at least 2.33; at the 5 per cent, 1.96.

Several trends are apparent. The most consistent difference is found between drama and the discussion methods. The next most consistent difference is between talk and discussion. In both instances the discussion came out lower than either the talk or the drama series.

The difference is less pronounced between talk and the drama. In the Lord Clive series, there was practically no difference; in the series on African Sleeping Sickness, the difference is significant at the 10 per cent level and this difference is in favor of the drama. In the series on *Les Miserables*, the difference is significant at the 10 per cent level also, but this difference is in favor of the talk.

A possible explanation of the fact that the discussion in all cases tended to be less effective is that discussion is used less frequently in presenting children's programs. It is possible that children are not accustomed to hearing that type of program; therefore it is not as effective in holding their attention. It is also possible that the classroom scene was too distasteful and familiar to prove effective. The analysis presented next shows that capable students who do not use radio much show greater differences among methods than capable students who use the radio frequently. The former reason may be more nearly correct. It is also possible that the discussion method is psychologically more difficult for acquiring knowledge.

Another analysis is necessary when the subjects are divided into groups differing in important developmental characteristics. For example, how do results compare when we take a group of subjects high in general ability and high in the amount of use of radio and a group low in general ability and low in use of radio? When the entire group was

Topic	Method	Number	Score
	High in Ability, Low in Radio Group		
Lord Clive	Talk	23	3.78
	Discussion	23	3.69
	Drama	23	3.69
African Sleeping Sickness	Talk	22	3.27
	Discussion	22	3.45
	Drama	22	3.59
Les Miserables	Talk	23	1.65
	Discussion	23	2.00
	Drama	23	1.78
	High in Ability, High in Radio Group		
Lord Clive	Talk	19	3.84
	Discussion	19	3.89
	Drama		
African Sleeping Sickness	Talk	18	3.50
	Discussion	18	3.50
	Drama	18	3.56
Les Miserables	Talk	20	2.70
	Discussion	20	2.95
	Drama	20	2.75

divided into the four possible groups, that is (1) high in general ability, high in use of radio, (2) low in general ability, low in use of radio, (3) high in general ability, low in use of radio, and (4) low in general ability and high in use of radio, using an arbitrarily chosen reference point of 6.5 (grade-norm) for reading test scores, and 9 hours a week for the amount of time spent in listening to the radio, it was found that two of the groups had in the neighborhood of twenty cases in each of the three program methods and could therefore be used in the analysis of variance. Figures showing the groups, the number of subjects in each group, and the mean initial test score on the factual portion of the tests are given in the preceding tabulation:

In the next tabulation, the results of variance analyses for these two groups are given:

Subject		Degree of Freedom	Sum of Squares	Variance	Ratio
	High in Ability, Low in Radio				
	Methods	2	4.298	1.149	
Lord Clive	Within groups	66	427.652	6.479	.18
	Total	68			
	Methods	2	32.398	16.199	
African Sleeping Sickness	Within groups	63	211.632	3.359	4.82
	Total	65			
	Methods	2	63.39	31.695	
Les Miserables	Within groups	66	200.61	3.04	10.43
	Total	68			
	High in Ability, High in Radio				
	Methods	2	35.065	17.532	
Lord Clive	Within groups	54	248.315	4.598	3.81
	Total				
	Methods	2	3.721	1.860	
African Sleeping Sickness	Within groups	51	172.666	3.386	.55
	Total	53			
	Methods	2	6.4	3.2	
Les Miserables	Within groups	57	449.2	7.89	.41
	Total	59			

Two of the six ratios shown in the preceding tabulation are significant above or near the 1 per cent level of confidence; these two are in the series dealing with African Sleeping Sickness and *Les Miserables* in the group that is high in general ability and low in the use of radio. This leads to the conclusion that for infrequent radio listeners there is a difference in effectiveness between methods of presentation, while for those who listen to the radio quite frequently, one method of presentation is about as effective as another.

Applying the previously mentioned "t-test" to the differences between means within these two series, the talk and drama methods in both cases are significantly above the discussion method but both are equally effective. Figures showing these relationships are given in the following tabulation:

Difference Between Methods	Topic	"t"
Talk and discussion	African Sleeping Sickness	2.06
	Les Miserables	4.55
Drama and discussion	African Sleeping Sickness	3.04
	Les Miserables	2.27

To be significant at the 1 per cent level of confidence, the "t" score must equal or exceed 2.58; at the 5 per cent level it must be at least 1.96. Thus we see that in both the series on African Sleeping Sickness and the series on *Les Miserables*, the group that was high in ability and low in use of radio showed significant differences between two of the methods. The talk and the drama methods were about equal in effectiveness, but in both cases the discussion was less effective than either.

Several possibilities are offered as explanation of the fact that for those who are infrequent radio listeners there is a statistically significant difference in effectiveness among

the methods, while there is none for those who are frequent listeners. One reason may be that those who listen to the radio regularly have developed a technique for listening to the various types of programs and thus learn equally from each type. Still another reason may be that those who listen to the radio a greater number of hours have a greater interest in the program as a means of presentation; they learn from the program no matter what method is used to present the material, merely because it is presented over the radio.

In summarizing the findings on the factual information section of each test, we may say that, taking the groups as a whole, the talk and the drama presentations came out significantly on top in one set of material each, and were about equal in effectiveness in the remaining series. The discussion was less effective than either of the other two methods in all three sets of material. Dividing the groups into those who listen to the radio comparatively frequently and rather infrequently, the talk and the drama methods are again more effective for those who are infrequent radio listeners, while there seems to be little difference in effectiveness for those who are frequent listeners.

In analyzing the data relative to attitudes, it was found that there was little consistent difference in the effectiveness of one method over the other. In the test on Lord Clive the attitude concerned the significance of various factors underlying human behavior. The discussion showed evidence of a mature attitude in twice as many subjects as either the talk or the drama, even though a larger per cent of those in the discussion group evidenced maturity on the initial test. In this test also, the talk and the drama methods were almost identically effective. In the group with a mature response on the final test the percentages were as follows: 44 per cent for talk, 74 per cent for discussion, and 44 per cent for drama.

In the test on *Les Miserables,* the talk showed a considerably larger per cent of mature responses on the final test than either the discussion or the drama methods. The two remaining groups, the discussion and drama, were about equally effective. Of the group giving mature responses on the final test talk was most effective for 93 per cent, discussion 76 per cent, and drama 88 per cent.

Which of these types of programs is most enjoyed by the listener? Is the interest related positively, or negatively, or not at all to the amount gained from the programs? By a comparison of the mean rating of interest given each program, and by an analysis of variance and 't-test," it was found that in two series the drama was significantly more interesting than the other two methods and that in the third series, although the difference was not as significant, there was the same trend of interest. In all cases, the discussion fell below the level of the other two methods. It would seem then, on comparison of these facts with the figures given previously, that the forms which presented the content material most effectively were also the most interesting, and vice versa.

In summarizing the analysis of data, the following statements are significant:

1. For presentation of simple ideas to the entire group of subjects, the talk and the drama methods were significantly more effective than the discussion at or near the 1 per cent level of confidence in two of the three content areas. In the third content area, while the difference was not as significant, there was a tendency for the talk and the drama to be more influential.

2. Again taking the group as a whole, the most consistent difference between methods for any one of the content areas was between the drama and the discussion. The next most consistent difference was between the talk and the discussion methods. In all cases, the discussion was significantly lower than either the talk or the drama.

3. When the entire group was divided on the basis of amount of radio listening as well as general ability, the talk and the drama were significantly more effective for those who are infrequent radio listeners, while there was no significant difference in effectiveness for those who listen to the radio quite frequently.

4. For the teaching of attitudes, taking the group as a whole, no one method was consistently more effective than any other and of the three attitudes presented, the scientific attitude seemed to be the most difficult for the listeners to grasp.

5. In interest value, for the two content areas in which there were significant differences in interest, the drama was rated most

interesting method with the discussion being rated the least interesting of the three. In the third content area, all forms were about equally interesting.

The foregoing conclusions apply only to the programs used in this study. More experimental work must be done with all three methods to determine the limits of effectiveness for each. For example, in the discussion method employed in this study the general background is that of the classroom with teachers and pupils carrying on the discussion. It would be interesting to see what happens when a different background is used, when a summary is provided at various points throughout the broadcast and at the close, and when listeners are supplied with various aids to be used during the broadcast. The effectiveness of such variations of the discussion procedure must be investigated. Similar questions may be profitably raised for other types of radio presentations.

A COMPARISON OF THE ENGLISH VOCABULARY USED BY CHILDREN OF NON-AMERICAN ANCESTRY IN HAWAII BEFORE THEY REACH THE AGE OF SEVEN YEARS WITH THAT OF KINDERGARTEN CHILDREN IN CONTINENTAL UNITED STATES

MADORAH E. SMITH
University of Hawaii

In the course of a study (3) made of the language development of children in Hawaii, verbatim records were taken of the conversation of these children. It became possible by analysis of these records to determine the frequency of use of each word occurring therein and thus to compare the English vocabulary of preschool children in Hawaii with that of children of similar age in continental United States. That differences in vocabulary would be found was anticipated on account of the prevalence of languages other than English and of the peculiar English dialect which are spoken in Hawaii. This dialect has been described in *American Speech* by Reinecke and Tokimasu (2). They report the idioms and common word substitutions occurring in the dialect but do not give any numerical data as to the effect such substitutions have on the frequency of use of the actual words. This paper is a comparison of the words most commonly used by young children in Hawaii with those of high frequency on the Mainland as reported in the International Kindergarten Union's study (1).

The list of most common words used by young children in Hawaii was determined by tabulating all English words found in 1021 fifty-sentence samples of the conversation of little children from two to six years of age which had been recorded while the children were engaged in free play about their own homes or at kindergartens. The records were taken by university students of the same racial ancestry as that of the children observed and were used as the basis of a larger study of the language development of the preschool population. These children form a fair cross-section of the population of Hawaii with the exception of those of non-Portuguese Caucasian ancestry, for their parents represented every type of occupation carried on in the Islands. Six of the samplings consisted of twenty-five children at each age level from two to six years, all of the same racial ancestry, Hawaiian and part-Hawaiian, Portuguese, Chinese, Japanese, Korean and Filipino. These six samplings were of children residing in Honolulu, as was also another sample of one hundred children of various racial antecedents whose records were taken while they were at kindergarten. Another sampling of 125 children was made of Filipino children in rural districts, mostly on the Island of Maui but some from Oahu. The remaining cases were about half of different races in Honolulu and the others, children of Japanese or Hawaiian descent on the island of Maui and in rural Oahu. The children's speech was recorded exactly as spoken, but the pronunciation was ignored.

The proportion of English words in the fifty-sentence samples ranged from an average of fifty percent for the children of Japanese ancestry to an average of ninety-nine percent for those of Portuguese ancestry; about seven-eighths of all the words recorded were English, half of the remainder Japanese, the rest in decreasing frequency from some Filipino dialect, Chinese, Hawaiian, Korean or Portuguese. Altogether a total of 142,574 English words, of which 2123 were different, were used by the children. All words used at least ten times by five or more different children were listed and compared with various other vocabulary lists. The frequency of other words of high frequency in these other lists was also determined from the Hawaiian records. This more complete list with the comparisons has been mimeographed and a limited number of copies are available (4).

The most important comparison made was that of the vocabulary of the children in Hawaii with that of children entering first grade published by the International Kindergarten Union (1). As for that list 489,555 running words were examined, the frequencies quoted in the published list were divided by 3.4 so that they might be directly comparable to those found in Hawaii. For consideration in this paper only those words are used which occurred either three times as often in the Hawaiian records as they did in the International Kindergarten Union's list after this adjustment had been made; or, although used at least twenty times in the latter study (i. e. six times after the division by 3.4 to make the lists comparable), were nevertheless either not found in the Hawaiian list or, if found, occurred only one-sixth as often as on the Mainland. Regularly formed inflected forms of words are combined and counted as one word in the comparison.

Since the children in Hawaii averaged 48.2 months in age, they probably average at least one year younger than the Mainland kindergartners. However, the effect of age difference is not so great as would appear; for, since the older children used longer sentences and a higher proportion of English, their records weight the word-frequency counts much more heavily than do those of the younger children. The average increase from two to six years was seven percent in proportion of English words and the sentences used were almost twice as long at six as at two years.

When the Hawaiian records were compared with a previous study (5) of the speech of Mainland children exposed to only one language, English, it was found that the children in Hawaii showed much less maturity in their speech, using shorter sentences, fewer connectives and pronouns, more interjections and making much less use of inflected and conjugated forms of words.

This lack of maturity of speech is shown at several points when the Hawaiian vocabulary is compared with that of the International Kindergarten Union's list (hereafter referred to as the Mainland list). There are sixteen interjections among the words used at least three times as often in Hawaii. These show a total frequency of 2700 in the Hawaiian list but only 188* in the Mainland list (see Table I–A). Of these, three words, "eh," "huh" and "yeah," are used in the Islands at the end of sentences declarative in form to turn them into questions instead of using such phrases as "isn't it," "doesn't he" or "won't you." These words occurred 804 times in the Hawaii list, 11 in the Mainland one.

On the other hand, among the words used less than one-sixth as often in Hawaii, eight pronouns are found used there 420 times as against 4376 times on the Mainland (see Table II–A). For four of these, "it," "its," "their" and "they" which occur 380 times in Hawaii as compared with 3946 times on the Mainland, a word " 'em" or "um" which may be a mispronunciation of "them" or "him," being used for both, is often substituted. Twenty connectives occurred only 1430 times in Hawaiian records but 10,311 in the Mainland ones.

The failure to inflect words found so often among young children and in pidgin English is shown in the fact that twenty-nine irregular past forms of verbs, four variants of the verb "to be," three third person singular forms, two past participles and the future auxiliaries in full or contracted forms are in the list of words used less than one-sixth as often in Hawaii as on the Mainland (see Table III). To these may be added three superlative forms, one comparative and two plural nouns, the total of whose frequencies in Hawaii is five; on the Mainland 150.

A peculiarity of the English dialect in Hawaii is a failure to use contracted forms. Six words combined with contracted forms of the verb "to be," four with the future auxiliaries, ten other verbs combined with "n't" and eight miscellaneous contractions are found among the under-used words in Hawaii, occurring a total of only 482 times, while on the Mainland they occurred 4289 times.

Other contrasts between the two lists are traceable to the local dialect. These are shown with explanatory notes in the under-used list in Table IV and in the over-used list in Table V.

As there were children of two years included in the Hawaiian survey, it is not surprising that two distinctly baby words "bye" or "bye-bye" and "wow-wow" were used much more, 123 times, in Hawaii than on the Mainland, two times. It is perhaps also

* All figures hereafter quoted for the Mainland study are those obtained after the original figures have been divided by 3.4 to make the frequencies in the two studies comparable.

due in part to these younger children that we find the more frequent use of words ending in a diminutive suffix as "dolly" and "horsie."

It often has been commented upon that when immigrants meet with a new language slang and swear words are among the first learned. The little children in Hawaii used many more such words than did the children on the Mainland. The difference may be due to a more extensive knowledge of such words among the immigrants' children or a greater concern on the part of the kindergarten teachers for the use of approved words by their charges. Since nine-tenths of the Hawaiian records were taken while the children were at play about their homes, usually with no other adult present than the recorder; and their mothers often were not aware that certain of the words their children used are considered undesirable, it is not surprising that such words occur more frequently in the Hawaiian list. If words primarily used when calling names, highly colloquial words, slang and swear words are all included, we find twenty-four such words used 1018 times in Hawaii and but 41 times on the Mainland as contrasted with but four slang and colloquial words in the under-used list with frequencies of 22 in Hawaii and 275 on the Mainland (see Table I–C and II–C.) Differences in the two sets of slang words are partly due to the time element between the two recordings—e. g. the word "scram" was not current in 1928.

Of course many of the differences in the lists are due to differences in environment. Although the majority of the children in both studies lived in the city, yet, owing perhaps to the plantation system prevailing in Hawaii and a difference in crops, it is probable that many of the things associated with farm life are more familiar to children on the Mainland than to those in Hawaii; or the words may have been taught to the Mainland children in kindergarten projects. Whatever may be the cause, we find in the list of under-used words such words as "corn," "farm," "hen," "lamb," "plow," "turkey," "w a g o n," "wheat" and similar words, all used a total of only 21 times in Hawaii but 527 on the Mainland.

Houses in Hawaii seldom are built of brick nor do they often have basements or chimneys. So only one of these words is used at all in Hawaii and that but twice, while on the Mainland the combined frequency was 84. Railroads are few and we find "engine," "station" and "track" occurring 14 times in the Hawaiian records to 108 in the Mainland ones.

The mild, even climate in Hawaii with very slight seasonal changes is pictured in the low frequency of such words as "coal," "fur," "Jack Frost," "sled," "stove," "summer," "winter" and "wool" which all together were used seven times in Hawaii, 424 times on the Mainland.

Plants and animals common on the Mainland but rare or absent in Hawaii are found in the under-used list, only 2 of the words being used at all and that but 13 times, while on the Mainland they were used 352 times (see Table II–H).

As the majority of the children's parents were immigrants and over half from non-Christian lands, the American holidays are not of so much importance in Hawaii, so we find "Christmas," "Easter," "reindeer," "Santa," "Santa Claus" and "Thanksgiving" used a total of 50 times in contrast to 511 times on the Mainland. Four words frequent in social intercourse are also under-used—"good morning," "ma'am," "party" and "visit," 10 to 108 times.

But the children in Hawaii picture their environment in their much greater use of the words: "cane," "coconut," "guava," "mango," "papaia," "pineapple" and "rice," 175 times to the Mainland children's 9. The proper nouns or adjectives "Chinese," "Hawaiian," "Japanese" and "Portuguese" used 73 times (Mainland 0) picture a race consciousness not matched by the Mainland use of "negro" 10 times (Hawaii 0). Neither group used America or American 10 times.

The different settings in which the records were taken is pictured in the comparative frequencies in the three lists of words associated with games, kindergarten and home (see Table I–F and G, II–K and L). There are two words under games used less frequently in Hawaii. Tag is played under a different name, but why "kite" should not have occurred more often is not clear. In the Hawaiian home list the references to relationship terms and such words are a picture of the setting in which the record was made.

In studying the Hawaiian, Portuguese and Korean words used by the children in Hawaii

TABLE I

WORDS USED AT LEAST THREE TIMES AS OFTEN BY THE CHILDREN IN HAWAII AS BY THOSE ON THE MAINLAND

A. Interjections

	H.#	M.#
ah	1025	12
aw	21	7
bang	67	10
boom	21	6
ah+	395	0
ha	66	18
hey	627	103
ho	19	0
huh+	75	11
meow	12	3
ouch	24	5
ow	24	0
uh	27	0
umhm	12	0
yah	51	13
yeah+	334	0

B. Baby Words

bye	107	2
horsie	42	4
wow wow	16	0

C. Name Calling, Slang and Swearing

	H.	M.
baldhead*	17	0
bogieman	13	0
boss	15	5
bullshirt or bulliar	20	0
crook	29	0
crybaby	17	0
damn	14	0
dumb	20	0
fellow	19	3
flop	15	2
fresh	16	4
gang	26	0
goofy	27	0
Jew	12	0
junk	28	2
lick	111	0
licking	51	4
lie	22	2
O.K.	228	0
pal	23	0
saucy	21	0
scram	69	0
shut (often in shut up)	211	15
sucker	10	4

D. Peoples and Products More Common in Hawaii

	H.	M.
cane	14	2
coconut	26	3
guava	14	0
mango	44	0
papaia	18	0
pineapple	18	4
rice	41	0
Chinese	21	0
Hawaiian	12	0
Japanese	30	0
Portuguese	10	0

E. Words With Special Reason for High Frequency

airship	16	0
candy	275	69
radio	30	10

F. Words Associated With Home

	H.	M.
auntie	166	45
bathe	21	6
clothes	161	39
daddy	519	136
gum	48	5
hungry	28	8
kiss	28	7
ma	491	5
mama	1670	128
match	29	3
medicine	17	3
mommie	26	0
mummy	20	0
mop	11	0
pa & papa	218	31
pants	52	16
pray	11	0
shirt	11	3
sleep	199	62
slipper	37	6
sonny	73	0
tobacco	10	0
toilet	20	5
towel	16	4
uncle	85	25

Frequency H. in Hawaiian word list, M. in International Kindergarten Union's List.
\+ All used at the end of declarative sentences to turn them into questions.
* Literal translation of a Japanese term.

TABLE I—Continued

TABLE I—Continued

WORDS USED AT LEAST THREE TIMES AS OFTEN BY THE CHILDREN IN HAWAII AS BY THOSE ON THE MAINLAND

G. Words Associated With Games	H.	M.	H. Verbs	H.	M.	I. Nouns	H.	M.	J. Modifiers	H.	M.
ace'	19	0	bite	134	18	blood	12	4	downtown	61	4
agate	69	0	buy	257	52	bracelet	10	3	last	34	10
bicycle	70	18	change	33	12	cent	63	21	long	65	11
boxing	13	2	crack	37	4	fair	42	8	lucky	14	0
card'	63	10	cry	115	20	folks	55	2	naughty	12	4
cowboy	34	7	fall	260	74	hook	21	4	quick	123	12
deal'	14	0	follow	26	3	ink	10	0	ripe	10	2
donkey'	22	6	give	1278	247	Kress	14	0	sour	16	4
fight	120	29	mind	30	6	noise	20	1	tight	21	7
goal	11	0	poke	53	0	peanut	20	4	tired	61	12
hide	94	18	push	227	41	puppy	51	10			
hit	148	32	scold	46	0	pussy	50	13			
jack	12	4	scratch	16	5	rubbish	18	0			
kick	36	12	smash	11	0	soda	13	2			
master	39	7	spill	14	3	stone	54	11			
pistol	11	0	teach	17	3	tire	28	7			
punch	52	11	touch	72	19	toe	21	5			
race	29	9				veranda	13	0			
safe	17	0									
spin	19	3									
throw	221	25									
win	56	7									

' Used almost entirely by children of Chinese ancestry when playing cards.

TABLE II

WORDS USED AT LEAST SIX TIMES BY MAINLAND CHILDREN AND EITHER UNKNOWN OR USED ONE-SIXTH OR LESS AS OFTEN BY THE CHILDREN IN HAWAII

A. Pronouns

	H.	M.
each	0	29
else	6	39
it	258	2948
its	4	97
their	5	86
these	18	211
they	113	815
those	16	151

B. Connectives

about	26	185
above	0	15
across	1	64
after	19	101
against	0	8
along	0	48
also	0	12
and	573	3452
at	104	700
because and cause	50	292
beside	0	9
both	0	37
either	0	14
into	2	67
maybe	6	55
of	59	1119
so	56	371
to	533	3709
upon	1	34
without	0	19

C. Slang

gosh	0	10
kid	21	70
looky	0	19
lots (of)	1	176

D. Associated With Farm Life

	H.	M.
barn	0	54
churn	0	11
corn	0	39
cotton	2	18
farm	2	35
farmer	1	37
geese	0	8
hen	0	14
lamb	1	6
plow	0	21
rooster	0	11
sheep	5	33
turkey	1	89
wagon	9	81
wheat	0	36
wheelbarrow	0	34

E. Houses

basement	0	10
brick	0	18
chimney	2	56

F. Railroad

engine	8	50
station	1	14
track	5	44

G. Climate

	H.	M.
cape	0	43
coal	2	25
fur	0	18
Jack Frost	0	15
melt	0	11
sled	0	29
snow	0	124
spring	1	30
stove	4	32
summer	0	49
winter	0	37
wool	0	11

H. Animals and Plants Rare or Absent in Hawaii

blackbird	0	16
bluebird	0	36
dandelion	0	19
deer	8	66
jack-in-the-pulpit	0	11
lily	0	6
oak	0	6
redbreast	0	9
robin	0	71
snake	2	11
squirrel	0	38
tulip	3	19
woodpecker	0	34

I. Holidays

Christmas	20	167
Easter	6	61
reindeer	0	28
Santa	0	43
Santa Claus	24	165
Thanksgiving	0	47

J. Social

	H.	M.
good morning	0	22
ma'am	0	14
party	10	64
visit	0	8

K. Associated With Kindergarten

bead	0	27
brownie	0	19
chain	4	29
circle	0	25
clap	0	6
clay	1	31
crayon	0	47
desk	1	9
fourteen	1	8
Indian	9	51
paint	13	101
paste	15	170
pattern	0	11
peg	0	11
round	11	79
row	2	13
scissors	7	78
square	0	25

L. Games

kite	5	52
tag	0	7

TABLE II—Continued

WORDS USED AT LEAST SIX TIMES BY MAINLAND CHILDREN AND EITHER UNKNOWN OR USED ONE-SIXTH OR LESS AS OFTEN BY THE CHILDREN IN HAWAII

M. Food and Eating

	H.	M.
breakfast	0	19
cheese	1	14
cherry	1	9
corn	0	39
dinner	4	27
dish	11	65
food	1	21
fork	0	10
gingerbread	0	15
jar	1	6
lettuce	1	14
onion	3	21
pancake	1	8
pear	0	9
peas	0	9
pie	0	38
popcorn	1	6
pumpkin	12	62
radish	1	19
salt	1	10
sandwich	0	9
strawberry	0	13
supper	3	22
tomato	2	12
vegetable	0	14
walnut	0	6

N. Clothing

apron	2	15
lace	0	8
nightgown	1	7
romper	0	18
suit	4	37
umbrella	1	20

O. Parts of Body

arm	2	27
breast	0	8
foot	3	28

P. Family

cousin	6	32
family	1	11
grandmother	1	27
grandpa	1	14

Q. Historical

buggy	0	22
cart	0	6

R. Associated With Time

afternoon	2	34
ago	1	22
always	6	55
day	35	178
ever	1	16
May	0	31
month	1	13
names of the days of the week	24	155
often	0	6
once	9	52
sometimes	12	76

S. Color Names

	H.	M.
black	40	119
blue	27	160
brown	10	87
gray	0	16
green	23	132
pink	4	38
purple	0	49
red	43	218
white	26	144
yellow	9	107

T. Associated With Space

bottom	0	20
end	1	21
middle	4	28
spot	0	7

U. Meteorological

clouds	1	13
star	4	45
sun	9	54
sunshine	1	9

V. People

child	0	11
fireman	0	32
firemen	0	9
nurse	0	9
postman	1	8
tailor	0	21
woman	1	14

W. Abstract and Indefinite

	H.	M.
anything	9	57
cost	1	12
everything	9	59
sign	2	11
sound	0	11
war	1	10

X. Abstract Verbs

answer	1	8
belong	2	12
drown	0	14
fit	2	14
guess	2	141
hope	0	7
listen	4	24
ought	1	36
remember	3	23
seem	0	8
suppose	3	41
surprise	0	11
wonder	1	24

Y. Other Animals

alligator	1	13
bird	26	173
camel	0	13
den	0	6
feather	3	20
lion	3	19
nest	0	81
pigeon	4	21
sparrow	0	19
wolf	1	12
worm	0	21

TABLE II—Continued

WORDS USED AT LEAST SIX TIMES BY MAINLAND CHILDREN AND EITHER UNKNOWN OR USED ONE-SIXTH OR LESS AS OFTEN BY THE CHILDREN IN HAWAII

Z. Concrete Verbs	H.	M.	A′. Nouns	H.	M.		H.	M.	B′. Modifiers	H.	M.
bend	0	8	alley	0	10	basket	8	184	anyway	2	13
bounce	1	13	auto	1	28	cupboard	0	5	asleep	0	8
build	3	88	city	0	8	curtain	0	11	awful	0	16
bump	0	18	country	4	43	dresser	0	16	beautiful	0	10
fill	1	19	fence	4	26	handle	2	45	dark	4	22
fold	2	31	flag	5	65	pail	1	6	electric	2	21
knock	7	42	garage	5	38	pan	0	14	few	0	8
slip	1	14	hall	0	9	rake	0	24	past	1	12
trim	0	11	lake	0	20	rug	3	19	poor	2	12
			land	0	9	shovel	2	16	quiet	2	12
			mill	0	15	stamp	1	14	quite	0	9
			mud	3	22	trunk	0	6	really	1	23
			New York	0	13	window	9	122	smooth	1	11
			ocean	1	9	board	6	59	such	0	20
			steam	0	14	bunch	2	96	together	0	41
			steeple	0	6	plant	12	71	warm	1	36
			street	9	75	pole	1	11	whole	12	72
			tent	0	13	stem	0	19	wide	0	10
			wall	3	21	wood	6	75	wooden	0	9

TABLE III

WORD FORMS USED SELDOM BECAUSE OF FAILURE TO CONJUGATE OR TO MAKE USE OF CONTRACTIONS

	H.	M.		H.	M.		H.	M.		H.	M.
am	80	540	made	29	410	aren't	1	39	wasn't	1	31
are	56	871	met	0	9	couldn't	1	86	we'd	3	46
blew	0	10	might	3	53	didn't	23	237	we're	2	117
bought	3	51	rang	1	7	doesn't	5	39	weren't	0	7
brought	8	58	rode	0	11	hadn't	1	17	we've	0	22
built	0	14	sang	1	15	hasn't	1	38	who's	1	14
came	35	166	sat	2	16	haven't	0	66	wouldn't	1	39
could	34	219	saw	39	683	he'd or he'll	0	12	you'd	1	8
did	33	405	seen	3	66	here's	4	170	you'll	1	24
does	1	71	sent	2	13	he's	14	77	you've	0	10
done	2	93	shall	3	70	I'd	0	40			
fed	1	8	should	3	69	I'll	29	285			
fell	8	54	sold	0	30	isn't	2	119	**Inflected Forms**		
flew, flies	1	18	took	16	138	it's	16	296			
goes	11	96	tried	1	18	I've	1	132	best	4	36
had	59	532	was	109	803	mine's	3	27	biggest	1	11
has	36	401	were	3	184	she'd or she'll	0	13	geese	0	6
hung	0	6	will	17	477	there's	10	109	longer	0	8
kept	0	9	wore	0	7	they'll	0	14	most	0	8
left	2	25	would	3	124	they're	1	44	stories	0	81

TABLE IV

WORDS USED AT LEAST SIX TIMES BY MAINLAND KINDERGARTNERS AND EITHER UNKNOWN OR USED ONE-SIXTH OR LESS AS OFTEN BY THE CHILDREN IN HAWAII. LESS FREQUENT USE DUE IN PART TO PECULIARITIES OF THE LOCAL DIALECT

	H.	M.			H.	M.
a or an	288	6338	"one" or "the" used instead when the article is not omitted altogether.	the	3035	4628
				one	1483	1175
almost	18	82	"Little more" used instead.			
afraid	6	38	"scare" used instead.	scare	29	8
cute	4	110	"smart" takes its place in part.	smart	36	0
dad	1	13	there appears to be a tendency to add "ie" to words more frequently.	daddy	519	136
doll	30	190		dolly	74	37
do	162	666	"make" used instead.	make	1027	1283
east, west, north, south	0	32	directions given by using place names and two Hawaiian words meaning to the "sea" and "mountains."			
have	198	1323	"get" often used instead.	get	1509	845
hunt	0	11	"find" used instead.	find	124	66
kitten	1	10	"pussy" used instead.	pussy	50	13
is	367	2281	"stay" or "get" used instead.	stay	419	78
just	52	281	"only" used instead.	only	297	83
	H.	M.			H.	M.
great	2	69	"big" is usually used instead of any of these three terms, also used in the phrase "come big" instead of "grow."			
large	0	5				
tall	7	18		big	411	229
grow	11	86				
little	115	996	"small" is much more commonly used; "little" is used mainly when pronounced "lili" in the phrase "lili more" for "almost," "hardly" or "nearly."			
tiny	3	19		small	141	33
wee	0	6		more	1313	255
hardly	0	9				
nearly	2	12				
outdoors	0	9	"outside" takes its place.	outside	173	38
pony	1	17	"horsie" is used instead.	horsie	42	4
porch	3	30	either the Hawaiian word *"lanai"* or "veranda" is used.	veranda	13	0
soon	3	38	another word for which "lili more" is used.			
through	13	91	two words which have been almost replaced by a Hawaiian word, *"pau"*			
finish	26	48				
tooth	0	9	apparently known only in the plural form.	teeth	37	32
wish	8	61	"like" is used instead.	like	2382	701
people	11	118	"folks" is used instead.	folks	55	2
stairs	0	12	"steps" is used instead.	step	49	62

to determine what words lingered longest when the ancestral language was disappearing before another tongue, it was found that words referring to food and eating, clothing, parts of the body and family lingered longest. As some of the children in Hawaii were using other languages to express such concepts and some foods or clothing were not used very often, we find many such words in the under-used lists (see Table II–M and N). Rice, however, is such a common food that it is found 41 times in the Hawaiian records, but it is not in the Kindergarten Union list. The only part of the body in the over-used list is "toe," perhaps because the children run barefoot all year long and most of them do not even own a pair of shoes.

Despite the fact that the children used their parents' languages for many family relationships, we find more such words in the over-used (Table I–F) than under-used (Table II–P) lists. That is partly because the Hawaiian records were taken at home but also because "ma" is Chinese as well as English and "mama" and "daddy" or "papa" are easier terms than those used in some of the other languages spoken in the Islands. It is interesting that the words for grandparents are in the under-used list. For these relatives who, if resident in Hawaii, spoke less English than the parents, the children used the terms in the ancestral language.

A few words suggest the passage of time. The Kindergarten Union list was published in 1928, the Hawaii study in 1939. So we find "buggy" and "cart" used 28 times on the Mainland, not at all in Hawaii; but

TABLE V

WORDS USED AT LEAST THREE TIMES AS OFTEN BY THE CHILDREN IN HAWAII AS BY THOSE ON THE MAINLAND. THE HIGHER FREQUENCY BEING DUE IN PART TO PECULIARITIES OF THE LOCAL DIALECT

	H.	M.	
already	203	60	Often used merely to express past time.
'ass (that's)	357	0	As in " 'ass mine" for "that's mine."
broke	367	37	Generally used in this form, break occurring only 11 times. Regularly used for "tear" or "torn."
by-'n-by	334	not listed	Often used to express simple future.
catch	255	42	Also used instead of "take" or "get."
chance	38	2	Every time used instead of "turn" in "taking turns."
care	119	26	In "no care" for "I don't care."
'em or um	1150	0	Used indiscriminately for every third person pronoun, especially instead of "it."
fool	30	2	Usually with meaning of "bother" or "meddle with."
go	4685	621	Used as a sort of universal auxiliary with other verbs to express past, future or infinitive.
going	2069	133	Used as in "going to" to express future but "to" is generally omitted.
kind	714	80	Used as the Chinese classifier is in phrases like "funny kind man"; in no case with the meaning "benevolent."
lazy	20	4	Used instead of "tired of" or "disinclined to do."
lend	158	0	Instead of "let" as in "Lend me look."
like	2382	701	Includes meaning of "want to," "wish" and "desire."
long	65	11	In "long time" used for "used to", "long ago" or "once."
me	3258	907	Often erroneously for other first person pronouns; by the Japanese children with a Japanese ending for "my" and with "I" as in the sentence "Me I have no pencil."
mines	56	0	For "mine."
more	1313	255	With "no" instead of "none" or "no," with "little" instead of "soon" or "almost," and with comparatives as in "more better."
no	4234	412	Instead of "not" or "n't" as in " no can."
only	297	83	Used for "just."
open	179	55	Includes idea of "unbutton" and of turning on light or power.
plenty	204	0	Used without "of" instead of "many," "greatly," or "very."
scare	29	8	Instead of "am afraid" as in "I scare."
shame	34	0	Includes meanings of "embarrassed" and "shy."
small	141	33	Used also instead of "little," "tiny" and "wee."
smart	36	0	Used instead of "clever," "bright" or "cute."
sore	199	7	Used instead of "ache," "hurt," or "pain."
stay	419	78	Used instead of many forms of the verb "to be" both as a copula or auxiliary as "He stay sleep." "Papa no stay" for "Papa is not at home."
tall	554	156	Used instead of "say."
try	198	51	Used with almost any imperative as in "try come here."
wait	272	57	Includes the idea of a threat instead of "watch out."
waste	118	5	In phrase "waste time" to express disinclination instead of "I don't want to."
wild	30	10	Used instead of "angry" or "scold" as in "Mama wild me."
yet	133	54	Frequently added to sentences to express or emphasize the past tense.

"radio" used 30 times in Hawaii, only 10 times on the Mainland.

The extensive use of two words in Hawaii was due to special circumstances. "Airship" (16 to 0) occurred only among the children at one of the Hawaiian kindergartens where they had been taught a song about an airship. "Candy" 's high frequency (275 to 69) was due to the Filipino group whose recorder was in the habit of loosening his small subjects' tongues by a supply of candies.

The greater maturity of the Mainland children's speech is shown by their greater use of words expressing time (632 to 91), of nouns indefinite and abstract (160 to 22), of meteorological terms (121 to 15) and of verbs of an abstract nature (356 to 19) (see Table II–R to X). They also used ten color names a total of 1070 times to the Hawaiian children's 182 (this may be due to kindergarten experience) and four space terms, "bottom," "middle," "end" and "spot," 76 times while the Hawaiian children used only two of the four "end," and "middle," and that but five times.

In the under-used list are 11 other words connected with animal life. Probably the Mainland children had had more experience

with circuses and zoos and may have been taught some of these words in kindergarten.

For a few of the remaining words in the over-used list (see Table I–H, I, J) reasons for their higher frequency may be suggested. Four words, "cent," "Kress," "soda" and "buy," suggest more frequent experience with shopping. Ten of the verbs and two nouns express violence or disaster. Probably the children in Hawaii were rougher at play outdoors than the others in kindergarten indoors.

For the remaining words in the under-used list, no explanation is offered. With the army so frequently on manoeuvers and living on an island, we would hardly have expected to find "tent" and "ocean" in this list. Other nouns suggest a difference in home furnishings, but the absence of many of these words from the vocabulary of the children in Hawaii indicates that marked hampering of expression of ideas in English must be felt when the children enter school.

To summarize: marked differences are found as to the particular words occurring in the vocabularies of children of Portuguese and non-Caucasian ancestry in Hawaii and of kindergarten attendants on the Mainland. These differences are shown to be due to the English dialect or pidgin English prevalent in the Islands, the use of words from languages other than English, differences in climate and environment, differences in the settings in which the records were made and in the comparative maturity in the use of language of the groups studied.

References

1. International Kindergarten Union. *A Study of the Vocabulary of Children before Entering the First Grade* (Washington, D. C.: International Kindergarten Union, 1928), 36 pp.
2. Reinecke, J. E. and Tokimasu, A. "The English dialect of Hawaii," *American Speech* (1934), 9, 48–58, 122–131.
3. Smith, Madorah E. "Some light on the problem of bilingualism as found from a study of the progress in mastery of English among preschool children of non-American ancestry in Hawaii," *Genet. Psychol. Monog.*, (1939), 21, 119–284.
4. Smith, Madorah E. *The English vocabulary used by non-"haole" children in Hawaii before they reach the age of seven years.* (A few mimeographed copies of the complete study available.)
5. Smith, Madorah E. "A study of some factors influencing the development of the sentence in preschool children," *J. Genet. Psychol.*, (1935), 46, 182–212.

FACTORS AFFECTING AGREEMENT IN TEACHERS' BEHAVIOR RATINGS OF NURSERY SCHOOL CHILDREN*

KATHERINE H. READ
Institute of Child Welfare
University of California

The use of behavior ratings in the study of personality raises questions as to which traits can be judged with most agreement by raters, and what factors may influence the degree of agreement. These questions can be considered to advantage in the nursery school because (1) the nursery school provides opportunity for extensive observation of the personality of each child, (2) the training and interest of the nursery school teacher centers on the personality development of the child and (3) several nursery school teachers are generally available to serve as judges or raters for each child.

In the present study, behavior ratings on 31 nursery school children have been analyzed to find the agreement with which judges rated certain traits and to discover possible factors influencing agreement on the ratings.

The thirty-one children rated in the study were members of three groups attending the Purdue University Nursery School. They were probably typical of children attending a University nursery school which charges tuition; they came largely from professional families, and ranged in age from 2 to 4½ years.

Each child was rated by from four to seven judges who were regular staff members or graduate students teaching in the nursery school. These judges had known each child that they rated for one semester or more. The groups were divided as follows:

Group A, consisting of eight four-year olds, was rated by six judges.

Group B, consisting of twelve three-and four-year olds, was rated by four judges.

Group C, consisting of eleven two-year olds was rated by seven judges.

* Material for this study was collected at the Purdue University Nursery School. The writer wishes to express her appreciation to the teachers in the Purdue Nursery School for their assistance in making the behavior ratings and to Dr. Herbert S. Conrad of the Institute of Child Welfare for his helpful criticism during the preparation of the manuscript.

As two of the judges rated children in two groups, and one judge rated children in three groups, the total number of individuals who rated was thirteen.

The rating scale used in the study was an abbreviated form of Conrad's California Behavior Inventory for Nursery School Children.[1] It contains items of behavior observable in nursery school situations. Each item is rated on a seven point scale with the levels one, four and seven defined briefly in terms of illustrative behavior. The abbreviated form includes 67 items of behavior selected as most significant by a group of experienced teachers and clinicians.[2]

With each group of raters preliminary training in the use of the rating scale was given by practice ratings made on at least ten traits not included in the abbreviated scale. The rating was followed by two or more meetings of the group of raters to check and discuss the practice ratings in order that the directions and trait definitions might be clearly understood.

The judges recorded their ratings for each trait on a separate blank containing the names of the children. The completed ratings were combined and tabulated with one blank for each trait.

In order to determine which traits were judged with the most agreement (without making an elaborate statistical study which the small numbers hardly seemed to justify), the difference between the highest and the lowest ratings received by each child on a trait was tabulated, and the average of these extreme differences was taken as indicating the amount of disagreement of the judges as to child-behavior on that trait. For example, on the trait *Popularity* there is an average difference of 1.9 intervals between the highest and lowest judgments of raters on children in Group A, as follows (each "x" in this table represents a single judge's rating):

TABLE 1

Name of Trait: Popularity

Child	Levels of Trait 1	2	3	4	5	6	7	Difference between highest and lowest rating
1.				xx xx	x x	x		2
2.			x	xx xx	xx			2
3.				xx xx	xx x			1
4.		x x	xx xx	x				2
5.		xx x	xx x	x				2
6.				xx x	xx xx			2
7.				xx x	xx xx			1
8.			x	xxx xx	x			2
9.		x	xx x	xx x				2
10.		x	xx	xx x	x			3
11.				xx	xx	xx x		2
Average difference								1.9

The average differences between extreme ratings on traits were listed in order for the three groups from the least average difference in judgment to the greatest average difference in judgment. This list is given in Table 2.

Table 2 thus indicates the average amounts of disagreement within each group of raters. For group A with six judges the average difference in extreme judgments on a trait ran from 1.1 to 3.75 intervals; for Group B with four judges the average extreme differ-

TABLE 2

Rank	Group A: Average difference of extreme ratings	Group A: Number of traits with this difference	Group B: Average difference of extreme ratings	Group B: Number of traits with this difference	Group C: Average difference of extreme ratings	Group C: Number of traits with this difference
1	1.1	2	.7	1	.9	1
2	1.5	2	.8	1	1.4	2
3	1.6	3	1	5	1.45	3
4	1.75	6	1.1	10	1.55	1
5	1.85	7	1.2	4	1.6	3
6	2	8	1.25	3	1.7	3
7	2.1	6	1.3	4	1.8	2
8	2.25	10	1.4	2	1.9	10
9	2.35	5	1.5	10	2	4
10	2.5	4	1.6	6	2.1	9
11	2.6	4	1.7	6	2.2	11
12	2.75	4	1.75	7	2.3	6
13	3	2	1.9	3	2.4	1
14	3.1	1	2	1	2.45	6
15	3.25	2	2.1	1	2.55	1
16	3.75	1	2.2	1	2.7	2
17			2.25	1	2.8	1
18			2.5	1	2.9	1
	Total	67	Total	67	Total	67

ence was from .7 to 2.5 intervals; and for Group C with seven judges the average extreme difference was from .9 to 2.9 intervals. The spread of these averages shows that there were wide differences in the amount of agreement with which traits were rated.

The traits were then listed in order according to the amount of agreement with which they were rated by the judges. This was done by noting the rank of the average extreme difference in judgment on a trait for each of the three groups of judges and then taking the average of these ranks. For example, *Special Peculiarity of Facial Expression*, the trait judged with the most agreement, had a rank of one in all three groups, making an average rank of one. The trait *Self Sufficiency*, judged with the least agreement, had a rank of 9, 14 and 16, making an average rank of 13. The list of traits in order from those rated with the most agreement to those rated with the least agreement is given in Table 3.

TABLE 3

Name of Trait	Rank in Group A	B	C	Average Rank
1. Peculiarity of Facial Expression	1	1	1	1
2. Inspection of hurts	1	3	3	2.3
3. Quarrelsomeness	6	4	5	5
4. Restlessness	5	4	7	5.3
Boastfulness	6	5	5	5.3
Ease with which the child is emotionally affected	4	4	8	5.3
5. Nervous habits	3	3	11	5.7
Irritability	7	4	6	5.7
Fear in non-social situations	12	3	2	5.7
6. Fear in response to bullying	3	5	10	6
7. Delayed retaliation (revenge)	4	4	11	6.3
Ease of stimulation of anger	7	2	10	6.3
Emotional response when thwarted by an adult	8	4	7	6.3
8. Reaction to teasing	15	3	2	6.7
Companions in activities	11	3	6	6.7
9. Fault finding	10	7	4	7
Jealousy of partiality to another child	9	7	5	7
Sulking	2	10	9	7
10. Proneness to antipathies	5	7	10	7.3
Leadership	5	9	8	7.3
Popularity	5	9	8	7.3
11. Compliance with suggestions of children	4	13	6	7.7
12. Degree of emotional response	6	7	11	8
13. Response with thwarted by child	9	5	11	8.3
Indirectness response to failure	4	9	12	8.3
Sustained effort in occupation	2	13	10	8.3
Compliance with suggestions of adult	4	10	11	8.3
14. Suspiciousness of people	11	4	11	8.7
Behavior reaction to defeat	12	5	9	8.7
Social inclination toward group	6	10	10	8.7
Rights (self assertion)	7	11	8	8.7
Behavior response to difficulty	5	12	9	8.7
15. Adjustability to new situations	8	9	10	9
Enthusiasm	11	8	8	9
Blaming others	6	4	17	9
Indiscriminate retaliation	4	9	14	9
16. Indirectness response to fear	5	9	14	9.3
Apprehensiveness	7	11	10	9.3
Ruthlessness	6	4	18	9.3
Attacking others	4	4	11	9.3
Friendliness to children	9	11	8	9.3
Friendliness to adults	8	6	14	9.3
Compliance with regime	8	12	8	9.3
Persistence of interests	8	12	8	9.3
Number of interests	10	10	8	9.3

TABLE 3—Continued

Name of Trait	Rank in Group A	B	C	Average Rank
17. Desire for affection	7	8	14	9.7
Sense of responsibility in play	8	11	10	9.7
Group cooperation	8	10	11	9.7
Speed of decision	11	10	8	9.7
Curiosity	10	9	10	9.7
Ambition	5	12	12	9.7
18. Reaction to social pressure	8	9	13	10
19. Inhibition of emotions	11	9	11	10.3
Bullying	10	12	11	10.3
Stammering	6	12	13	10.3
20. Behavior reaction toward sympathy, approval adults	7	13	12	10.7
21. Dependence on outside direction	12	6	15	11
22. Talking to others	13	8	3	11.3
23. Social inclinations toward individuals	6	15	14	11.7
Self reliance	13	12	10	11.7
24. Affectionateness	11	9	16	12
25. Bossiness	8	17	12	12.3
Sense responsibility for group	14	11	12	12.3
Happiness	12	16	9	12.3
26. Recovery from emotional disturbance	16	11	11	12.7
Finality of decision	15	12	11	12.7
27. Self sufficiency	9	14	16	13

The traits at the beginning of the list are traits in which the judges agreed in their ratings of child behavior. On some of the traits as *Peculiarity of facial expression, Inspection of hurts, Quarrelsomeness, Restlessness, Boastfulness, Ease with which the child is emotionally affected,* and *Irritability* the agreement in judgment is close in all three groups of judges, as indicated by the agreement in rank of the trait in each group. One might consider that the traits in the first nine ranks are traits in which the judges showed relatively close agreement in their ratings of child behavior. All groups had a small spread in judgment on these traits.

The traits in the middle nine ranks are traits in which the agreement among judges is neither especially high nor especially low. The position of these traits in the list might very well shift under different circumstances such as with different judges etc.

The traits in the last nine ranks are traits in which all groups of judges showed relatively wide disagreement in rating child behavior.

What are the differences between all these traits?

As one would expect, the "overt" traits tend to be judged with most agreement. For example, the judges agreed in rating *Quarrelsomeness,* which involves overt behavior, while they differed widely in their judgments on the trait *Recovery from emotional disturbance,* which involves interpretation of behavior. *Ease with which the child is emotionally affected* was judged with agreement while *Emotional inhibition* was not: the child who is emotionally affected tends to cry or hit, but inhibition of emotion has no equally observable form of expression. Similarly, the judges agreed in rating *Boastfulness, Irritability, Restlessness,* while they tended to disagree on *Self Sufficiency* and *Finality of Decision.*

While traits easy to judge were "overt" traits, certain fairly overt traits appeared difficult to judge. There seemed to be a tendency for more disagreement over traits where the child's attitude or behavior toward adults was involved. The personality of the adult may play a part here and may actually bring out different behavior in the same child. For example, *Affectionateness, Dependence on outside direction, Reaction to sympathy or approval of adults,* even *Friendliness to adults* to some extent, were traits judged with disagreement in each group of judges. These traits seem characterized by the fact that behavior in response to some adult is concerned.

The question may be raised here as to the variability of a child's behavior. Lack of

agreement may be due to large variations in the child's behavior even in such a relatively constant environment as the nursery school. Certain behavior patterns such as *Restlessness, Fears* may be fairly well established; but social techniques like *Bossing* or *Bullying*, and other traits such as *Sense of responsibility for the group, Self reliance* and *Emotional inhibition*, may not have developed into consistent patterns for most children. This would affect agreement among raters. The large amount of disagreement in judging children on traits relating to social and emotional behavior may indicate real variation in child behavior.

Lack of agreement may also indicate traits or behavior to which the teacher gives relatively little attention and of which she is less aware when watching a child's behavior. In contrast to *Quarrelsomeness*, a trait such as *Happiness* may be difficult for a teacher to rate because she may give relatively less attention to this less compelling but certainly equally important trait.

The teacher may also fail to recognize the early manifestations of certain traits in the simple or unfamiliar form in which they appear at the nursery school age. The behavior of a child who goes up and snatches an unwanted toy from another child may not be considered under *Social inclinations toward individual children* by a teacher whose attention is centered on the poor social techniques shown by the child.

The further explanation may be offered that teachers' ratings are less likely to agree for behavior which is not directed toward the teachers, as *Bossing, Bullying*, assertion of *Rights*. The teachers, rarely being the recipients of this behavior, may give it less attention and thus be less likely to agree in rating it. The basis for disagreement here differs from that for traits previously noted in which the child's attitude toward the adult seemed to be a factor making for disagreement in rating.

The lack of agreement in judging *Stammering*, a relatively objective trait, probably reflects a difference in standard as to whether the general hesitation or repetition normal in the speech of young children was considered by the rater as *Stammering*.

The amount of variation shown in judging child behavior in this relatively favorable situation would seem to indicate that a large variety of factors influence ratings. These factors should be carefully considered when ratings are employed in the evaluation of an individual personality. It may be possible to obtain greater reliability in ratings (1) if traits are selected for rating that are likely to be judged with agreement, or (2) if judges give more attention to areas of behavior that are usually rated with less agreement.

Summary

Behavior ratings of 31 nursery school children have been analyzed to discover the agreement with which 67 traits were rated, and the factors which may have influenced agreement among raters.

The children were rated on an abbreviated form of Conrad's Behavior Inventory for Nursery School Children, consisting of 67 items of behavior selected as most significant by a group of experienced teachers and clinicians. In the Inventory each trait is rated on a seven point scale from most to least.

Every child was rated by from 4 to 7 judges who were teachers in the nursery school, either regular staff members or graduate student assistants who had known the children rated for one semester or more. There were thirteen raters in all.

The amount of disagreement among judges on a trait was determined by finding the extreme difference of the judges' ratings for each child, and averaging these extreme differences for all the children on each trait. The average difference in extreme ratings on a trait ran from 1.1 to 3.75 intervals in Group A (eight children rated by six judges), from .7 to 2.5 intervals in Group B (twelve children rated by four judges), and from .9 to 2.9 intervals in Group C, (eleven children rated by seven judges). This spread would seem to indicate fairly wide differences in the amount of agreement with which traits were rated.

The 67 traits were then listed in order of agreement of judges in rating, beginning with those traits in which judges agreed most closely down to those in which the amount of disagreement was greatest. Traits in the first third in order of spread of average difference of extreme ratings may be considered traits in which the judges showed relatively close agreement in their ratings of child behavior. Traits in the middle third in order of spread

are traits in which the agreement among judges was neither especially high nor especially low. Traits in the last third in order of spread of ratings are traits in which all groups of judges showed relatively wide disagreement in rating child behavior.

Factors which appeared to influence the agreement of judges included: (1) the extent to which the trait involved direct overt expression, (2) the extent to which the relationship with the rater was involved, (3) the extent to which the rater's attention was likely to be directed toward the behavior, (4) the extent to which the rater was able to recognize the manifestation of the behavior, and (5) the extent to which the child's behavior in respect to the trait was likely to vary.

Bibliography

1. Conrad, Herbert S. *The California Behavior Inventory for Nursery School Children*, (Berkeley, California, University of California Press, 1933).

2. Read, Katherine H. "Significant Characteristics of Preschool Children As Located in the Conrad Inventory," *Genetic Psychology Monographs*, 1940, 22, 455-487.

CHARACTERISTICS OF TEACHERS WHO ARE "LIKED BEST" AND "DISLIKED MOST"

ARTHUR T. JERSILD
Teachers College, Columbia University

I. ADULT RECOLLECTIONS OF ELEMENTARY SCHOOL TEACHERS

This paper reports findings in a more or less informal series of studies in which an effort was made to ascertain some of the qualities that characterize school teachers who are liked or disliked by their pupils. The information on which the studies are based was variously obtained by means of written reports, compositions, personal interviews, and, in a limited number of cases, by means of records of children's conversations with their parents or with each other.

The data consist in part of retrospective reports by adults, based upon childhood recollections, concerning the teachers whom they liked or disliked when they attended elementary school. Since this part of the investigation was undertaken first, the results obtained in it will be reported first in this paper.

The adults who supplied descriptions of their childhood teachers were members of one of the writer's classes at Teachers College. Mimeographed directions were distributed to each member. These directions simply asked the student to think back over his elementary school days, to try to recollect his various teachers, to single out the teacher whom he liked best, and the teacher he disliked most (or liked least), and, in each case, to describe in writing the characteristics and qualities of the teacher in question. It was expressly directed that the reports were to be anonymous. Other questions included in the directions called for information as to whether, in the judgment of the person submitting the report, the teachers who were described were similarly liked or disliked by other pupils at the time. Another section of the directions asked the students to describe an "ideal" elementary school teacher. The findings obtained in response to these additional questions are not reported here; the responses concerning recollections of the attitudes of other pupils tended to be rather indefinite, and the descriptions of the "ideal" teacher tended either largely to duplicate the accounts of the "best liked" teacher, or, in some cases, to combine concrete characterizations with general philosophical concerns.

The final scheme used in classifying the data is shown in Table 1; the categories were arrived at after several preliminary classifications. In the process of defining and refining the categories, the writer and a helper* independently analyzed and classified portions of the data in terms of tentative lists of categories. The judgments of the two workers were then compared, item by item, to find points of agreement and disagreement, and the categories were further defined, modified, contracted or expanded, to meet the difficulties and disagreements that were encountered. The final draft of categories, as shown in Tables 1 and 2 were also tested by the procedure of independent classification by two workers of a portion of the data. The agreement between the two workers was over 85 percent.

A glance at the categories under which the data are presented in Tables 1 and 2 will indicate that the scheme of classification is somewhat arbitrary, on at least two counts. First, even though it was possible, on the basis of the verbal definitions of the categories, and the wording of the reports that were classified, to make substantially accurate distinctions (as measured by agreement between independent analyzers) between the broad headings shown in the tables, it is not likely that the limited categories represent all the meanings intended by those who submitted the reports. It is likely, also, that items quite similar in wording might represent somewhat different meanings, and that items dissimilar in wording, and classified under separate headings, might represent substantially similar meanings. Secondly, it is obvious that varying configurations of personality factors cannot be described ade-

* The writer is grateful to Miss Ruth Herzmann for aid in this project.

quately in a brief report, such as that obtained from each subject in this study, or in a classification scheme such as that shown in the tables. As against this it may be said, however, first, that the reports that were submitted did uniformly stress qualities and characteristics, often in isolation, such as those represented in the tables, and secondly, that the classification scheme followed such descriptions as faithfully as possible.

It will be noted that the cumulative tallies exceed the number of individuals who reported. Most of the subjects reported more than one characteristic, and each item belong-

TABLE 1

CHARACTERISTICS MENTIONED IN REPORTS OF ADULT RECOLLECTIONS OF ELEMENTARY SCHOOL TEACHERS WHO WERE LIKED BEST. THE VALUES SHOW THE NUMBER OF PERSONS WHO NAMED EACH CHARACTERISTIC

	Women	Men	Total
N=	117	20	137
I. Characteristics relating to the teacher's personality and qualities as a human being			
Was kind, sympathetic, gentle, friendly, did not make children feel tense or afraid	52	6	58
Adjusted to individual children; was interested in, and liked, children as persons; showed insight into characteristics, needs and problems of individual pupils	43	7	50
Was vivacious, cheerful, outgoing, smiled easily, approachable as a person, not stiff or distant	42	4	46
Had sense of humor	19	7	26
Was adjustable, a good sport, no "front", not sensitive or on the defensive, put on no airs, ready to concede mistakes, etc.	4	3	7
Was tactful, socially well poised, well-bred, not vulgar	4	1	5
II. Characteristics relating to appearance, grooming, age, etc.			
Attractive physical appearance; good looking; had pretty eyes, etc.	36	2	38
Nice voice (specifically mentioned)	17	0	17
Nice clothes, well groomed (mentioned as a separate item or in conjunction with the two foregoing)	16	1	17
Youthful, not elderly, not an old maid	9	0	9
Good health	1	0	1
III. Characteristics relating to general conduct as disciplinarian and manager of order in classroom			
Was fair, just, had good sense of justice; did not punish wrong child, did not make all children suffer for shortcomings of one	12	5	17
Was impartial, had no pets or favorites	10	0	10
Discipline effective but not unreasonably strict; no discipline for its own sake, discipline firm but on basis of mutual respect	12	3	15
Was firm, assured, knew what he or she was about, standards high but consistent and predictable	9	3	12
Had good discipline, there was good order in classroom (apart from mention of nature of discipline imposed, as above)	6	0	6
No corporal punishment	1	0	1
Did not treat failure to learn as a moral wrong or as a form of misbehavior	2	0	2
Not a scolder, nagger, not fussy, was calm, even-tempered, patient	19	4	23
IV. Participation in children's everyday, or out-of-school, or outdoor interests, games, hobbies and activities			
Participated in children's outdoor activities and games, and in their extra-curricular interests	11	0	11
V. Characteristics as teacher, teaching; competence and effectiveness as a teacher			
Made things interesting, taught imaginatively, enthusiastic and resourceful as a teacher	28	5	33
Knew a lot, thorough as a teacher, we learned a lot from him or her, was well-informed, put things clearly, stressed what was important	16	3	19
VI. Miscellaneous			
School always tidy, neat (specific, limited to that particular point)	2	0	2
"She liked *me*," etc.; specific characterization of egocentric sort (as distinguished from general characterization as above)	25	3	28
Was intelligent	1	0	1

TABLE 2

CHARACTERISTICS MENTIONED IN REPORTS OF ADULT RECOLLECTIONS OF ELEMENTARY SCHOOL TEACHERS WHO WERE "DISLIKED MOST" OR "LIKED LEAST"

	No. of persons naming the characteristic		
	Women	Men	Total
N =	117	20	137
I. Characteristics relating to the teacher's personality			
Was unsympathetic, harsh, sarcastic, caused humiliation, anxiety, fear, was cruel, ridiculed children	50	7	57
Lacking in interest, insight, or adaptation as regards needs, interests and difficulties of individual children as persons; no liking for children as persons	25	2	27
Affected manner	1	0	1
Egocentric, constantly talked about self, talked about past prowess at football, etc.	2	0	2
Glum, sour, cross, seldom or never smiled, grouchy, abrupt, disagreeable	13	2	15
No sense of humor	7	2	9
Bad tempered, "terrific temper, lost temper"	9	2	11
Unapproachable, unable to adapt to child's level, stiff, cold, formal	9	0	9
Not a good sport, always right, never conceded self wrong, argumentative, sensitive	5	0	5
No refinement, vulgar, lacked dignity and breeding, became familiar with children, etc.	8	1	9
Fussy, nervous, fretted over trivialities, wrangled, irritable over small things, showed constant tension	20	3	23
Flighty, silly, scatter-brained, not level-headed	4	0	4
Emotionally unbalanced, muttered to self	3	0	3
No tact	1	0	1
Dishonest	1	0	1
II. Characteristics relating to appearance, grooming, etc.			
Unattractive physical appearance, ugly, "high color," extremely thin, etc.	13	0	13
Voice unpleasant, harsh, grating, rasping	13	1	14
Unattractive as to clothes and grooming, careless in appearance, always same dress, dull hair, etc.	12	1	13
Elderly, oldish, old-maidish	8	1	9
Poor health	3	0	3
III. Characteristics relating to general conduct as disciplinarian and manager of order in classroom			
Was unfair, punished innocent child, made whole class suffer for misconduct of a few children, etc.	7	5	12
Showed favoritism, was partial, had pets	17	1	18
Bossy, domineering, autocratic, overbearing	8	1	9
Discipline unnecessarily rigid, unreasonable, inflexible, stern objection to trivial infractions, discipline for its own sake	24	5	29
Was perfectionist in details	1	1	2
Was a constant scold, always nagging, always spoke in loud voice (as distinct from specific mention of unpleasant voice as such as above)	5	4	9
Not firm, inconsistent in discipline, didn't see things through, unpredictable as to discipline	10	2	12
Meek, spineless, timid	1	0	1
Discipline poor, no control, children paid no attention to her, poor executive ability	10	4	14
Used corporal punishment (or threatened to do so)	11	4	15
Treated and punished ignorance and failure to learn as a misdemeanor	6	1	7
IV. Participation in children's everyday, or outdoor, or out-of-school interests, games, hobbies, activities			
No participation in children's outside games, etc.	1	0	1
V. Characteristics as teacher, teaching competence and effectiveness as a teacher			
Didn't know much about subject, poor organization, unable to scale assignments, children didn't learn much, couldn't make self understood	9	2	11
Teaching was uninteresting, uninspired, unenthusiastic, cut and dried, stereotyped, teacher not interested in teaching	25	4	29
No judgment as to assignments, expected impossible amount of work	2	1	3
Indoctrinated children in own beliefs	3	0	3
Used unwise forms of motivation, gave material prizes, aroused too much competition, indiscriminate in praise and reproof	7	0	7
VI. Miscellaneous			
Classroom untidy, in disorder	2	0	2
Unintelligent	1	0	1

ing to a separate category received a separate entry. However, in no case was more than one tally entered for a given subject in a given category (e. g., "she wore awful clothes and her hair was done up in a funny way," received one entry, even though two details are mentioned, each of which, if appearing alone, would receive a tally under the clothes and grooming category).

An outstanding feature of both Tables 1 and 2 is the fact that a large proportion of the characteristics that are mentioned relate to the teacher's personality—his or her qualities and merits as a person or human being rather than specifically as a pedagogue. To be sure, many of the descriptions relate to the teacher's role as a disciplinarian and instructor, but it is likely that a large part of the table would be duplicated if the subjects had been asked to describe, not teachers, but associates, colleagues, employers, neighbors and acquaintances whom they "like best" and "like least." Practically all of the items in the first category, concerning characteristics of the teacher's personality and qualities as a human being, as well as many of the items in category II, represent generally desirable qualities rather than qualities peculiarly appropriate to the head of a classroom. To the items in these two categories may be added items such as fairness, firmness, and a sense of proportion, included in category III, which, while designating the teacher's conduct as a disciplinarian, also represent qualities that would be quite as favorably recognized in other situations. Only a relatively small number of items (little more than 22 percent of the total number of points raised in the characterizations of teachers liked best) refer to the enterprise of learning as such.

An incidental feature of Table 1 is the relatively high frequency of mention of matters concerning appearance and grooming. The category, "Attractive physical appearance; good looking, etc.," is mentioned by more people than is any category concerning the work of the teacher as an instructor. It appears, however, that there is a sex difference in emphasis on this category; it is mentioned by 36 of the 117 women but by only 2 of the 20 men.

Table 2, which deals with teachers who were disliked, roughly parallels Table 1. Again "human" rather than strictly "pedagogical" characteristics are mentioned most frequently. The main difference appears in the category concerning the teacher as a disciplinarian; unfavorable items in this category were mentioned relatively more frequently in descriptions of teachers who were disliked than were favorable items of this category in descriptions of teachers who were liked.

Relatively little mention is made of some of the more obvious little techniques which a teacher might use to participate in the everyday interests and preoccupations of the children. In the descriptions of teachers who were liked, only 11 persons mentioned items that fall in the category entitled, "Participation in children's everyday or out-of-school or outdoor interests, games, hobbies or activities." Failure to show interests of this sort is mentioned in only one description of teachers who were disliked. Even when items dealing with the teacher's ability to arouse and to tap the pupils' interests in connection with the work of the class are added, the total in this general category is still relatively small.

The foregoing discussion, and the tables on which it is based, suggests that in looking back over their school days, these people more or less took for granted the matter of teaching and learning and tended to regard their teachers primarily as persons who had to be lived with, listened to and looked at for several hours a day rather than as guides or mentors in the business of getting wisdom from lessons and books.

II. CHILDREN'S ACCOUNTS OF THEIR TEACHERS

ARTHUR T. JERSILD AND FRANCES B. HOLMES

This report is based upon responses obtained from pupils who were asked to describe the teachers whom they "liked best" and "disliked most" (or "liked least").

The subjects included elementary school and high school pupils. 526 children described teachers whom they liked, and 373 of these also were asked to describe teachers whom they disliked.* The grade and sex distribution of the children is shown in the tables that follow. The pupils represented several widely-separated localities, including

* The data in this section of the study, as in the preceding section, were procured through the collaboration of students in the writer's classes, whose aid is gratefully acknowledged.

communities in New Jersey, West Virginia, New York City, and the suburbs of New York. The children were variously approached in homes, in out-of-school groups, and where it seemed agreeable to all concerned, in schools.

It was emphatically directed that under no circumstances should an approach be made to groups or to individuals if the procedure or the results were likely to cause embarrassment to anyone. The reports by the children were obtained either in the form of brief compositions or in the form of replies to questions in individual interviews. In the directions given to those who cooperated in gathering the data, it was specified that the pupil should first be asked to "think about the teachers you have had since you started school"; then, "think about the teacher you like best . . . what was she like, why did you like her?" Similar questions were raised to elicit replies concerning teachers who were not liked. It was urged that in no case should the child be asked to reveal the name of the teacher, and, if the name were revealed, it was not to be recorded. As might be expected from the fact that children tend to think in concrete terms, the names of teachers often were divulged and, in the case of the groups that wrote brief compositions, the teachers' names not only were mentioned but became a part of the record, but no attention was given to the identity of the teachers in the treatment of the results. As it happens, the procedure of eliciting compositions from the pupils was used primarily only in obtaining information concerning teachers who were liked, and it may be added that the results obtained from compositions did not differ materially from the results obtained in private interviews.

In preliminary efforts to classify the characteristics and qualities mentioned by the children, use was made of the categories that were devised in the earlier analysis of the reports submitted by adults. The final scheme of classification conforms substantially to this list of categories, but includes certain changes and amplifications. To obtain a measure of the extent to which the categories were usable and objective, two independent workers analyzed and classified a substantial portion of the data, and then subsequently compared their judgments, item by item. The percentage of agreement in analyzing descriptions made by 55 children of teachers whom they liked best, including a total of 86 different descriptive items, was 92. In another comparison, involving accounts both of "likes" and "dislikes" by 41 children, including a total of 93 descriptive items, the agreement was 91 percent.

The categories are listed in Tables 3 and 4. As in the treatment of reports by adults, the classification of the material had to be somewhat arbitrary. In many instances, no doubt, children used different words to express similar meanings, or intended different meanings when they used similar terms. All that could be done was to take the child's own words and apparent intentions and classify them according to categories and definitions upon which independent workers showed a high degree of agreement. Undoubtedly in many instances a turn of phrase would place one child's report in one of the categories in Group I ('human qualities as a person'), while a different phrasing would put another child's report in Group III ('characteristics as a disciplinarian or as director of class') even though both children may have meant about the same thing. The groupings do not imply that a teacher embodies a number of discrete traits and that she can separately be analyzed as a human being, as a disciplinarian and as a pedagogue. Rather, the classification represents the aspects of personality or performance which the children happen to emphasize in the particular words they use in describing their teachers.

The description of each category included in the tables represents an abridgment of a lengthier descriptive account, with illustrations, that was used in classifying the data. Since the descriptions are relatively clear it has not been deemed necessary to reproduce lengthy lists of illustrations in the children's own words, but a few examples may be given to illustrate categories that are phrased in more general terms. Under the heading, "Kind, sympathetic, etc." came items such as: "she didn't scare us; didn't hurt our feelings; she was so generous; so motherly." On the negative side were such items as: "she is sarcastic; always making fun of you; she was mean." The group "Interested in, and liked pupils as persons" included such expressions as: "she seemed to like all of us; she was like an old friend; she talked to us like human beings; she seemed to know what to do for each one of us." On the negative side,

TABLE 3

CHARACTERISTICS MENTIONED BY CHILDREN IN DESCRIPTIONS OF TEACHERS WHOM THEY "LIKED BEST". THE VALUES SHOW THE PERCENT OF CHILDREN NAMING EACH CHARACTERISTIC

	Boys and girls combined			Boys (b)	Girls (b)	Boys	Girls
Age (a)	5–8	9–12	13–17	9–12	9–12	13+	13+
Grade	1–3	4–6	7–12	4–6	4–6	7–12	7–12
N	25	203	298	38	40	103	128
I. Human qualities as a person							
1. Kind, sympathetic, considerate	16	24	16	8	10	14	20
2. Interested in, liked and adjusted to pupils as persons	0	1	16	3	0	16	13
3. Natural, approachable, companionable, good sport	0	3	8	8	0	9	6
4. Vivacious, cheerful, smiling	8	3	9	3	13	6	8
5. Sense of humor, joked, funny	4	1	5	5	3	5	4
6. Good-tempered (not bad-tempered, cranky, etc.)	4	7	2	5	8	0	2
Dignified, "not fresh" or common	4	0	.03	0	0	0	0
II. Physical appearance, grooming, voice							
1. Attractive, good looking, pretty eyes, nice figure, etc.	0	6	5	3	13	5	8
2. Nice voice, nice manner of talking	0	2	2	5	0	3	1
3. Attractive clothes, well-groomed, neat, clean	0	3	2	3	3	0	4
4. Youthful, not elderly	0	.05	1	0	0	0	2
III. Characteristics as a disciplinarian or as director of class							
1. Fair, just	4	15	11	21	20	10	10
2. Impartial, no favorites or pets	0	5	5	3	3	4	5
3. Discipline effective, respected, consistent	4	11	16	21	13	13	14
4. Didn't scold, yell, holler, shout	12	5	8	8	8	8	8
5. Did not use corporal punishment	4	2	0	3	0	0	0
6. Did not treat ignorance as moral wrong	0	0	1	0	0	2	0
7. Lenient, easy-going, let us have own way	4	4	3	5	10	6	0
8. Did not require hard work or "too much work"	16	2	9	0	3	5	6
IV. Participation in activities; providing gifts, entertainment							
1. Joins in or permits games and play, tells stories, etc.	32	7	4	5	3	7	5
2. Special attentions: parties, food, trips, hobbies	8	4	2	5	5	1	2
V. Performance as teacher, teaching							
1. Interesting, resourceful, enthusiastic as teacher	4	4	13	8	0	14	13
2. Supplied interesting school projects, activities, experiments, etc.	24	12	5	3	20	5	5
3. Taught, explained, well; pupils learned much	8	26	34	26	18	31	34
4. Helped individual pupils with lessons	4	5	10	5	0	9	3
5. Permitted expression of opinion, voice in class affairs	0	2	2	3	0	1	2
VI. Miscellaneous and general							
1. Miscellaneous: he's a man; intelligent, room tidy, etc.	0	2	1	0	3	2	0
2. Egocentric; she liked *me;* praised *me;* gave *me* prize, etc.	20	17	5	8	20	6	6
3. Indefinite, general, no specific reason	12	4	.05	16	11	5	7

(a) The "age" entries represent ages of children for whom no grades were indicated; the main classification was according to grade.

(b) In some cases, the sex of the children was not indicated; such subjects were included in the results for boys and girls combined. This accounts for the fact that this total is larger than the sum of the separate boy and girl totals.

TABLE 4

CHARACTERISTICS MENTIONED BY CHILDREN IN DESCRIPTIONS OF TEACHERS WHOM THEY "LIKED LEAST". THE VALUES SHOW THE PERCENT OF CHILDREN NAMING EACH CHARACTERISTIC

	Boys and girls combined			Boys (b)	Girls (b)	Boys	Girls
Age (a)	5- 8	9-12	13-17	9-12	9-12	13+	13+
Grade	1- 3	4- 6	7-12	4- 6	4- 6	7-12	7-12
N	9	99	265	30	30	94	106
I. Human qualities as a person							
1. Unkind, unsympathetic, ridicules	11	9	16	7	7	13	18
2. Lacking in interest in needs, characteristics, etc. of pupils as persons	0	1	2	0	3	2	3
3. Not natural, unapproachable, prim, self-important, not good sport	0	2	7.3	0	0	8	8
4. Never smiled, sour, glum, solemn	11	0	4	0	0	4	4
5. No sense of humor	0	0	0	0	0	0	0
6. Cross, ill-tempered, cranky	11	19	18	10	27	30	13
7. Tactless, "fresh"	0	4	3	0	13	4	3
8. Nervous, jittery, fussy	0	1	1	3	0	0	1
9. Flighty, silly	0	0	1	0	0	0	1
10. Emotionally unbalanced, "queer," "talked to self"	0	0	2	0	0	2	3
11. Dishonest, "sneaky," told lies	0	0	0	0	0	0	0
II. Physical appearance, grooming, voice							
1. Unattractive of feature, face, figure, "ugly face", etc.	0	3	2	0	0	2	4
2. Unpleasant voice, too loud, rasping, too low	0	4	4	0	7	2	5
3. Unattractive clothes and grooming, "sloppy" clothes, unattractive hair	11	2	3	7	0	3	4
4. Elderly—poor health	0	4	2.3	7	0	2	4
5. Unpleasant mannerisms	0	0	1	0	0	0	1
III. Characteristics as disciplinarian or director of class							
1. Unfair, unjust, punished too severely or all for error of one, etc.	0	13	8	13	10	10	6
2. Partial, had pets	0	10	11	17	3	7	9
3. Discipline rigid, too strict, inconsistent, poor	22	26	21.3	27	20	18	19
4. Constant scold, yells at you	56	5	9	7	10	7	13
5. Used corporal punishment	0	15	3	3	10	4	2
6. Treated failure to learn as moral wrong, impatient with error	0	1	2	0	3	2	0
7. Too lenient	0	2	1	3	3	2	2
8. Preachy, bossy, domineering	0	1	4	3	0	8	0
9. Never praises	0	0	.3	0	0	0	1
IV. Participation in children's interests, entertainment, etc.							
1. Did not join in children's games, hobbies, interests, etc.	11	2	1	3	0	0	3
V. Performance as teacher, teaching							
1. Uninteresting, uninspired, dull, boring	0	1	4	0	3	5	2
2. Uninteresting assignments, cut and dried, never do interesting things	0	0	1	0	0	2	1
3. Not a good instructor, didn't know much, not clear, etc.	0	9	10	10	7	4	8
4. No help to individual pupils in lessons	0	2	3	0	3	2	2
5. No judgment as to assignments	0	2	0	3	0	0	0
6. Allowed no expression of opinion, choice	0	0	1	0	0	2	1
7. Standards too high in lessons	0	1	1	0	0	0	0
8. Indoctrinated pupils with own beliefs	0	0	1	0	0	2	0
9. Used unwise forms of motivation	0	0	0	0	0	0	0
10. Too much homework	11	14	14	7	10	8	12
VI. Miscellaneous and general							
1. He's a man; she's a woman; not intelligent; room untidy, poor order; because of subject	11	3	1.3	3	0	0	1
2. Egocentric,—she disliked *me*. gave *me* bad report	11	7	5	17	3	6	4
3. No reason given—indefinite	0	7	6	3	13	8	5

(a) See footnote (a), Table 3.
(b) See footnote (b), Table 3.

such items as: "she didn't seem to like anybody."

The "Fair, just" group included a number of instances in which a child simply said that the teacher was "fair" or "unfair," as well as many instances in which the child described what he deemed to be fair dealing by the teacher, or items such as: "punishes children for nothing; makes everyone stay in when one child is bad; says I talk and I don't; his marks aren't right."

On the negative side, the category "Not natural, approachable," includes such items as: "talked too much about herself; thinks she's pretty; thinks she owns the school; conceited." In the "Undignified, tactless" group: "a little too friendly; doesn't mind her own business;" "Unpleasant mannerisms" included items such as, "picks nose, waves hands all the time; spits in handkerchief."

The category entitled, "Discipline effective, respected, consistent" includes a variety of characterizations such as: "not too strict; gave us just enough freedom; when she yelled you deserved it; she made us work" as well as descriptions of instances in which the teacher handled a disciplinary problem in a calm and reasonable manner. On the negative side were items such as: "too strict; after us all the time; we have to keep so quiet all the time" (also includes objections to specific disciplinary methods, other than corporal punishment, such as "made us stay in during recess").

"Joins in or permits games and play, etc." includes an item such as "he roots for our side when we play baseball," but in a larger number of items mention is made of games initiated by the teacher in the classroom or permitted by the teacher in the gymnasium or on the playground. On the negative side: "she doesn't ever play with us." The category "Supplied interesting school projects, activities, experiments, etc." denoted activities undertaken in connection with a lesson or unit of study as distinguished from periods of free play. On the negative side were items such as, "cut and dried; we don't do enough things; never varies the work."

A variety of items is represented under the heading, "Taught, explained well; pupils learned much," including such accounts as: "she knew a lot; you learned a lot from her; she explains things good; taught me more than any other teacher; goes right to the point; taught reading swell; she would teach things not right away but gradually." On the negative side: "gives us work we don't understand; muddled; too much emphasis on one subject." The category immediately following this concerns help given by the teacher to individual pupils in connection with their lessons as distinguished from the more general qualifications just mentioned. In the "Indefinite, general" category are items such as the general remark "she's nice"; miscellaneous items such as: "she's a woman; I just like her; she could be worse; she's always in the room, etc."

In tabulating the data, as in the tabulation of adult recollections, a tally was entered for each category represented in the child's description of the teacher, and only one tally was entered for any given category. Separate entries were made, however, for items in the different categories listed under the large (Roman numeral) groupings in Tables 3 and 4. The number of tallied items does not correspond to the number of children, since some would list characteristics belonging to several different categories, while some listed only one or two items.

Tables 3 and 4 show results in terms of age or grade levels, as well as separate results for boys and girls at grade levels four to six and seven to twelve.

A number of interesting trends appear when a comparison is made between the characteristics mentioned by adults in recollections of their elementary school teachers and the characteristics mentioned by children at various age levels. Table 5 summarizes, under six general headings, the results that have been shown in more detail in the preceding tables. Each value in the table represents a percentage, obtained by dividing the number of items falling into a given category by the total number of items or characteristics mentioned by the subjects at each of the various age levels.

It should be pointed out that two factors complicate the comparison of results based upon adult recollections and the results based upon reports by the children: First, the adult group consists largely of women, while the children represent an approximately equal number of boys and girls. This complication can be offset to some extent, however, by taking account of the results separately shown for males and females in the more detailed

TABLE 5

GENERAL SUMMARY, IN TERMS OF SIX LARGE CATEGORIES, OF RELATIVE FREQUENCY OF MENTION OF VARIOUS CHARACTERISTICS IN DESCRIPTIONS BY PUPILS OF TEACHERS WHOM THEY "LIKED BEST" AND "LIKED LEAST OR DISLIKED MOST" AND IN SIMILAR DESCRIPTIONS BY ADULTS ON THE BASIS OF RECOLLECTIONS OF THEIR ELEMENTARY SCHOOL TEACHERS

	Teachers liked best				Teachers disliked most			
Age level of persons reporting	5– 8	9–12	13–17	Over 20	5– 8	9–12	13–17	Over 20
Grade level of persons reporting	1– 3	4– 6	7–12	Post-graduate	1– 3	4– 6	7–12	Post-graduate
Number of persons reporting	25	303	298	137	9	99	265	137
Total number of items mentioned	49	370	604	443	15	170	459	414
I. Human qualities as a person	18	22	28	43	20	21	32	43
II. Physical appearance, grooming, etc.	0	7	5	18	7	8	7	13
III. Qualities and performance as a disciplinarian	23	24	26	19	47	43	35	31
IV. Participation in pupils' activities, games, non-academic interests	20.5	6	3	2	7	1	1	.2
V. Performance as a teacher, teaching	20.5	27	32	12	7	17	18	13
VI. Miscellaneous	18	14	6	7	13	10	7	.7

earlier tables. Second, adults were asked to confine their reports to recollections of elementary school teachers, while the reports from children include a number at the high school level. It still is possible, of course, to compare the adult results with results obtained from children at grade levels one to three and four to six.

There are outstanding differences between the trends in the reports by adults and by children. Adults emphasize general personality traits, as distinguished from specific characteristics of her performance as a teacher, considerably more than do the children. The category, "human qualities as a person," accounts for 43 percent of the descriptions given by adults, as contrasted with from 18 to 32 percent of the characterizations reported by the children. The children, on the other hand, give relatively more emphasis to qualities and characteristics of the teacher's performance as a teacher. At each age level, the children make more frequent mention than do adults of the teacher's qualities and performance as a disciplinarian, including such features as fairness, impartiality, reasonableness and consistency. In like manner, especially in descriptions of teachers who were liked, the children give considerably more stress than do adults to the teacher's competence as a teacher, her ability to present subject matter in an interesting manner and her effectiveness in helping the pupils to learn their lessons.

Adults mention items concerned with the teacher's physical appearance, dress, grooming and quality of voice considerably more often than do the children, but the difference here appears to represent a sex rather than an age difference. The detailed table of the reports submitted by adults shows that 20 percent of all items mentioned by women in their recollections of teachers whom they liked best were in this category, but only 5 percent of the items mentioned by the men were in this category. Likewise, it appears from Tables 3 and 4 that girls mentioned appearance, dress, grooming and voice quality more often than did the boys.

In the results based upon reports by children of teachers whom they liked best, it appears that the characterizations that are mentioned fall largely into the three categories, "human qualities as a person;" "qualities as a disciplinarian or director of order in the classroom;" and "performance as a teacher, teaching;" with approximately an equal distribution of frequencies between these three categories. There is an increase with age in the percentage of replies in each of these categories. This increase occurs mainly at the expense of the group of characteristics entitled, "participation in pupil activities and provision of gifts, games and entertainment." The latter category accounts for one-fifth of the items mentioned by the youngest children, and only three percent of the items mentioned by the oldest.

The oldest children place considerably more emphasis on characteristics relating to "performance as a teacher, teaching" than do the youngest children. As shown in Table 3, not only is there a drift with age toward greater frequency of mentions in this large classifica-

tion, but there is a change with age also in the specific features of the teacher's performance that are mentioned. The older children, considerably more frequently than the younger ones, emphasize items such as, "she explained well, we learned a lot from him, she knows a lot, etc." The younger ones place greater stress on the teacher's provision of interesting projects and experiments. The older children, apparently, place more emphasis on the teacher's ability to help them to acquire knowledge as something of value in itself, while the younger ones give more emphasis to the interesting qualities of the projects or undertakings by means of which the teacher endeavors to stimulate learning. This trend is interesting, even though, of course, it may not represent a genuine change with age in the pupil's judgment as to what is most valuable.

There is a notable difference between the reports of children and adults with respect to emphasis upon the teacher's ability to make lessons interesting and to help the pupils to master their lessons. Only 12 percent of the characteristics mentioned by adults in recollections of the teachers they liked best dealt with items in this category, while from 20 to 32 percent of the items mentioned by children at the various age levels dealt with items in this category.

A distinct age trend is exhibited in connection with the category that includes items of an egocentric character, such as, "she liked me; she gave me prizes," etc. In descriptions of teachers who were liked best, the percentage of items represented in this category is 20 in grades one to three and 5 in grades seven to twelve. In the adult reports, 6 percent of the items fall in this category. In descriptions of teachers who were disliked, there likewise is a decline with age in the data obtained from children (from 11 to 5 percent) in the frequency of items in this category. The adult accounts of teachers who were disliked included no items in this category. It would appear from this that pupils' judgments tend to become more impersonal as the pupils grow older, but the difference may spring less from a genuine decline in egocentricity than from a change in the manner in which personal attitudes are phrased.

On many points there is a difference between the qualities emphasized in descriptions of teachers who were liked and of teachers who were not liked. In the children's reports of teachers who were disliked, objections to the teacher's qualities as a disciplinarian occurred relatively much more frequently than did commendations on the same score in descriptions of teachers who were liked. The teacher's participation in children's activities, and provision of interesting pastimes is mentioned more frequently in descriptions of teachers who were liked than is the absence of these provisions in descriptions of teachers who were not liked. Personal appearance, dress, grooming and voice quality are mentioned somewhat more often in descriptions of teachers who were disliked than in accounts of teachers who were liked.

The category, "performance as a teacher, teaching," including ability or lack of ability to make things interesting and to help the children to master their lessons, is mentioned considerably more often in descriptions of teachers who were liked than in descriptions of teachers who were not liked. Within this general category there likewise is a shift in relative emphasis: there is frequent mention of homework requirements in descriptions of teachers who were not liked, but no mention of this in descriptions of teachers who were liked.

By way of a supplement to the data reported above, records were obtained of the spontaneous comments made by one child—an eleven-year-old boy—concerning his teacher during a period of twelve weeks. Unknown to the child, these comments were noted and recorded by his mother, who also had occasion to converse with many of the boy's classmates concerning their reaction to the same teacher.

The boy commented about his teacher on sixteen different occasions. As it happened, the teacher in this case was not popular with this boy (who preferred his teacher of the preceding year) or with his classmates, so that unfavorable comments predominated. Even so, his sentiments varied from time to time and favorable comments were made on six of the sixteen occasions. Following is a summary of the reasons he gave for liking or disliking the teacher; the number of items exceeds sixteen by virtue of the fact that on some occasions he gave more than one reason for his attitude:

Favorable	Times mentioned
"Nice to me"	2
"Likes me"	1
"Lets us have a good time"	1
In a good mood, doesn't yell	1
Marks fairly	1
Interested in class games	1

Unfavorable	Times mentioned
Unfair in punishments	6
"Favors the girls"	3
"Cross, mean, sore"	3
"Doesn't keep promises"	1
"Makes work uninteresting"	1
"Preaches and talks too much"	1
"Has moods"	1

As indicated above, there were sharp fluctuations in the boy's attitude. In the first record of his comments he complains that the teacher broke two promises (to allow children to paint, to include the boy's cartoon in the class paper), that "she punishes you for things you don't do," that "she talks so much and repeats things a million times." Two weeks later, he makes this comment: "Gee, the teacher was nice today. The two sixth grades played each other in football and she cheered and yelled and was so anxious for us to win." In his next comment he confesses to his mother that the teacher that day had caught him and a classmate in the act of passing notes, yet the lengthy account which he gives is highly favorable to the teacher: she was "awful nice." The note, intercepted by the teacher, had contained "bad words" which the boy would not repeat to his mother because they were not "fit for a woman's ears." But instead of showing anger when she talked to the boy and his classmate after school, the teacher "talked so nice to us . . . said she was disappointed . . . she wasn't going to punish us and we said we'd never do anything like that again . . ." The boys' account indicates that he was quite impressed by his own guilt, and by the teacher's handling of the situation and he vowed to do something nice for her.

The next record includes charges that the teacher was "so mean," followed, two days later, by a report that "school was swell today" (the boy and a classmate had tidied the school library "from force of habit" and "that sure got us in good with the teacher—she was pleased and praised us"). A few days later he reports the teacher is "hot and bothered," followed at later times by varying reports to the effect that she is better than the substitute teacher, that she had praised his drawing ability, that she had unfairly kept some of the boys in the classroom during gym period to write numbers backward from 500 to 1 (. . . "she said if we were so silly, she'd give us something silly to do"); that "that teacher gives me a pain;" that she had marked very fairly on a new type of report card (she had taken the boy aside and had explained his marks); that "I'm fed up", —the teacher had put his name on the board "just because I talked in assembly;" that "all of us boys are so sore . . . none of the girls ever miss gym . . . it's a gyp." The final record gives an account of the boy's report of how he, with a delegation of other boys, went to the teacher and in a "nice" way protested that she was partial to the girls; the teacher apparently did not welcome this protest.

Many features of the above account suggest interesting leads for further study. Although the child's attitude toward his teacher is predominantly unfavorable, it appears from the fact that he comments only sixteen times in twelve weeks that his feelings about his teacher tend to be quite episodic. As far as can be judged from his conversation, it appears that during a good deal of the time he is conscious neither of a distinct like or dislike for the teacher, but takes her as a matter of course. It is interesting also to note the happenings that bring a sharp increase in his enthusiasm for the teacher: her use of a quiet reprimand when he expected punishment, her show of eagerness for her pupils to win a game, a brief word of praise, etc. Although the child has many misgivings about his teacher, he seems hungry for a bit of praise or recognition, and he swells with pleasure when a morsel falls his way. It likewise appears in several reports of his grievances that the boy seems tacitly to recognize that he has misbehaved; he seems to object less to the fact that discipline is imposed than to the nature of the punishment. A similar note is discernible on the occassion when he seemed most keenly to feel that his teacher was "awful nice:" he seemed to be convinced of his own guilt and he was struck by the fact that the teacher simply talked to him and did not scold or punish as he had expected. Expressions of this sort, both on the positive and negative

side, appeared frequently in the reports obtained in the larger survey. Items to the effect that the teacher was strict but fair and, so to speak, made the punishment fit the crime, were frequent in accounts of teachers who were liked, just as items to the opposite effect often appeared in descriptions of teachers who were disliked.

Summary

This study deals with the characteristics of teachers who are "liked best" and of teachers who are "liked least" or "disliked most." It is based upon (a) reports written by adults who were asked to look back upon their elementary school years and describe, as best they could, the qualities of the teachers whom they liked or disliked; and (b) written and oral reports submitted by elementary and high school children. All written reports were anonymous and the subjects were asked not to name the teachers in question if they could avoid doing so.

The adults were graduate students, including 117 women and 20 men. The children represented pupils in grades 1 through 12; 526 described teachers whom they liked and 373 of these also were asked to describe teachers whom they disliked.

After several preliminary classifications, the characteristics described by the subjects were classified according to certain major and a number of minor categories. These categories are listed in Tables 1–4. The agreement between independent workers in classifying the data according to these categories ranged from 85 to over 90 percent.

The categories that were employed in treating the data are not used, of course, with the implication that a teacher embodies a number of discrete traits; rather, the classification grew out of the reports submitted by the subjects and it represents the characteristics and aspects of personality and performance which the subjects happened to emphasize.

A detailed summary of the findings appears in Tables 1 to 4 and a more general summary appears in Table 5.

A large proportion (43%) of the characteristics named by adults in describing childhood teachers whom they liked or disliked fell in the general category "Human qualities as a person." This category includes items such as kindliness, liking for others, vivacity, sense of humor, etc., and the opposite, as distinguished from qualities uniquely related to the office of teaching. A substantial but considerably smaller percentage of the characterizations named by children also fell in this general category.

Characteristics that could be classified under the general category "Qualities and performance as a disciplinarian" (including strictness and laxity, fairness and unfairness, partiality and impartiality, etc.) were mentioned more often by the children than by adults. Further, there were considerably more frequent mentions of negative qualities in this general category, in descriptions of teachers who were disliked, than of positive qualities in this category in descriptions of teachers who were liked.

Characteristics falling under the general category "Performance as a teacher, teaching" (including descriptions of the teacher's ability to make things interesting and clear, to help the pupils to understand and to "learn a lot," etc.) were mentioned considerably more often by children than by adults in descriptions of teachers who were liked best. Moreover, there was an increase with age in emphasis upon items in this category (from 20.5% in grades 1–3 to 32% in grades 7–12, as compared with 12% in the case of adults). This category was mentioned less frequently both by adults and children in describing teachers who were disliked than in describing teachers who were liked.

Adults emphasized details of "Physical appearance, dress, grooming, etc." considerably more than did the children in descriptions both of teachers who were liked and of teachers who were disliked. Items in this category were mentioned relatively more frequently by women than by men but the number of men included in the study was too small to provide a conclusive comparison. Physical attractiveness was also mentioned more frequently by girls than by boys.

Children in grades 1–3 emphasized items in the general category "Participated in pupils' activities, games, non-academic interests" considerably more than did the adults or children in the grades beyond the third.

Data gathered by a mother, who, unknown to her child, kept a record of the child's remarks at home concerning his teacher over

a period of several months, indicate that there may be decided fluctuations from time to time in a pupil's attitude toward his teacher. Since these records are limited to one child they are quite inconclusive but they suggest that further data would provide significant information concerning teacher practices that arouse favorable or hostile reactions in pupils.

In general it may be said that adults in appraising their childhood teachers tend largely to judge these teachers in terms of qualities which usually are deemed to be desirable in human beings in any walk of life as distinguished from qualties relating peculiarly to the work of a teacher. On the other hand, children, who still have the experience of living with their teachers from day to day, phrase their characterizations of teachers more in terms of specific features of the teacher's performance as the head of a class and as a pedagogue.

FACTORS IN THE PERSONALITY OF NURSERY SCHOOL CHILDREN

T. W. RICHARDS
The Samuel S. Fels Research Institute
Antioch College

In 1939 Ball and Roberts[1] published certain data regarding the Merrill–Palmer Personality Rating Scales, largely material concerning their reliability and the standards for their use. They reported also a table of intercorrelations between the scales. Since very little has been done with factorial analysis at the nursery school level, and the work of this sort in the area of personality variables at any age level is in little agreement, I thought it would be interesting to see what factor pattern this table of intercorrelations would yield.

Van Alstyne[3] reported at least three factors present in ratings of child behavior in 33 situations: "Pupil desirability," "Leadership," "Social Consciousness." Possibly factors of "Emotional Security" and "Responsibility" are also present. Williams[4] found two factors significant in 30 ratings of traits of social behavior in nursery school children, which scales were developed by Berne[2]: approach—withdrawal and ascendance — submission. Each of these studies employed older interpretations of the centroid method.

An added reason why this possibility was interesting was the fact that the Merrill–Palmer Rating Scales are a unique type of scale; constructed like the Thurstone Attitude Scales (items weighted by judges' opinions are summated for the score on each trait), the subject does not fill in his own blank (as is the case with the Thurstone Scales) but adult observers simply indicate whether the child behaves as per each item on each scale.

The table of intercorrelations, taken from the report of Ball and Roberts[1] appears in Table 1. Using Thurstone's centroid method three factors seemed significant. Variation in successive matrices of residuals appears in Table 2. Table 3 presents the factor loadings, unrotated and rotated, and the communalities for each factor.

In naming the factors, eight judges (on the psychological staff of the Fels Research Institute) were presented with the scale

TABLE 1

MATRIX OF INTERCORRELATION BETWEEN MERRILL–PALMER PERSONALITY SCALES, TAKEN FROM ROBERTS & BALL[1]

	Soc.	Asc.	Face Reality	Resp. Author.	Compl. Routine	Resp. Prop. R	Indep. Adult Aff.	Attractive. Person.	Phys.
Sociability		.10	.56	.45	.40	.45	.37	.70	.35
Ascendance			—.07	—.03	.—10	.17	.20	.17	.08
Face Reality				.62	.62	.45	.50	.49	.23
Resp. Author.					.75	.74	.31	.44	.13
Compl. Routine						.74	.28	.28	.25
Resp. Prop. Rt.							.37	.35	.55
Indep. Adult Aff.								—.25	.17
Personal Attrac.									.43
Physical Attrac.									

TABLE 2

SIGMAS OF COEFFICIENTS IN SUCCESSIVE MATRICES, MERRILL–PALMER PERSONALITY SCALES

Matrix	Sigma of Residuals
Correlation	.248
1st residuals	.098
2nd residuals	.117
3rd residuals	.074
4th residuals	.039
5th residuals	.032

σ mean, correlation matrix = .092

TABLE 3

UNROTATED AND ROTATED LOADINGS AND COMMUNALITIES MERRILL-PALMER PERSONALITY SCALES

	I	II	III	k_1	k_2	k_3	h^2
Compliance w. routine	.88	.03	.27	.725	.512	.030	.79
Resp. to authority	.86	—.02	.38	.759	.467	.146	.82
Resp. property rights	.67	.50	.34	.832	.154	—.311	.81
Tend. face reality	.52	.08	.60	.734	.094	.245	.61
Indep. adult affect	.36	.52	.02	.447	—.011	—.393	.36
Sociability	.14	.27	.86	.745	—.372	.312	.69
Phys. Attractiveness	.09	.20	.58	.500	—.274	—.202	.36
Attractiveness person	.07	—.05	.92	.604	—.299	.590	.80
Ascendance	—.15	.42	.09	.131	—.350	—.230	.19

names and their factor loadings. Names given by these judges appear in the following tabulation:

Judge	Factor I	Factor II	Factor III
1	Ascendance	Conformity	Likeableness
2	Resourcefulness	Compliance	Agreeableness
3	Objectivity of Interest	Self-Discipline	Sociality
4	Dominance-Aggressiveness	Submissiveness	Social Maturity, Extroversion
5	Self-Sufficiency	Docility	Sociability
6	Ascendance, Aggressiveness	Conformity	(related to) extroversion, or Likeableness
7	Independence, Self-Assertion	Submissiveness, Social Compliance	Social Attractiveness
8	Independence, Social Maturity, (non) goat, Ascendance, Originality, Self Assurance	Cooperation, Submissiveness, Tractibility Useful Citizenship	Popularity, Likeableness, "Goodness," Sheep, Halo.

Certainly a large number of trait names appear in the results furnished by these judges, but I think there are threads of continuity throughout sufficiently to call the traits "Self-sufficiency," "Conformity," and "Likeableness" (or halo, which is probably likeableness anyway).

REFERENCES

1. Ball, Rachel Stutsman and Roberts, Katherine Elliott. "A study of personality in young children by means of a series of rating scales," Journal of Genetic Psychology, (1938), *52*, 79–149.
2. Berne, Esther Van Cleave. "An experimental investigation of social behavior patterns in young children," *University of Iowa Studies in Child Welfare, 4,* #3, pp. 93.
3. Van Alstyne, Dorothy. "A new scale for rating behavior and attitudes in the elementary school." *J. Ed. Psychol.*, (1936), *27*, 677–693.
4. Williams, Harold M. "A factor analysis of Berne's 'Social Behavior Patterns in Young Children.'" *J. Exper. Educ.*,(1935).

READING READINESS AND READING ACHIEVEMENT IN FIRST GRADE

SYDNEY ROSLOW
Board of Education, Hastings-on-Hudson

INTRODUCTION

Recent studies have generally shown that placement in first grade is not advisable solely on the basis of a chronological age of six years. For example, Betts[1] reports that investigators have found that a mental age of six to six and one-half is essential for successful reading achievement. Hildreth[2] states that children below the range from 6 yrs. 3 mos. to 6 yrs. 6 mos. do not succeed as well as older children. On the other hand Gates[3] has presented data to indicate that children with a mental age of five can successfully be taught to read when the individual child's abilities are considered and adequate methods and materials are utilized. Furthermore, it has also been shown that children under six with above average intelligence can make satisfactory progress in first grade.[4]

In addition to the factors of age and intelligence, other factors are undoubtedly related to reading success in the first grade. These factors have been incorporated into the concept of reading readiness. These factors imply that the child has had a variety of visual, motor, auditory, and language (verbal) experience all related to reading. They include: acquiring a fund of information about objects and events common to childhood; building and making things; playing cooperative games; singing songs and reciting rhymes; listening to stories being told and read; telling stories and exploring pictures; drawing, tracing, painting, cutting, and pasting; conversing, talking, acting, and playing language games; etc.

Tests have been designed to measure the development of reading readiness and these tests have been proven to be positively related to reading achievement. Wright[5] found a correlation .44 and .61 between the Metropolitan Readiness Tests and the Gates Primary Reading Tests and teacher's marks respectively. Monroe[6] reported for her reading aptitude tests a correlation of .75 with reading achievement measured by Gray's Oral Paragraphs and the Iota Word Test. Yet from Monroe's manual it is also brought out that even though children may be deficient in reading aptitude, they may attain average success in reading. Twenty-five percent of those who scored at the 20th percentile on the aptitude tests became average readers. Similarly, Grant[7] wrote that "a high readiness score alone does not always assure success in reading; and a low readiness score does not always mean failure."

Since children are nevertheless placed in first grade when they are six years of age, the problem facing schools is one of insuring reading success for as many children as possible; with whatever degree of intelligence or of reading readiness they possess. It is the problem of preventing the reported average failure of 20 to 25 percent of first grade children principally because of reading difficulty.[8]

PROBLEM AND METHOD

The problem in this investigation was to use measures of intelligence and reading readiness in the placement of children into one of three first grade sections and then to compare the relative reading achievement of these three sections at the end of the first year. The purpose of this placement procedure was to assure as high a proportion of successes as possible. It was believed that homogeneous grouping with respect to ability and readiness would make for more effective teaching. In addition a program of remedial treatment and prevention was executed. Chil-

[1] E. A. Betts. The Prevention and Correction of Reading Difficulties. Evanston, Row, Peterson and Co., 1936, p. 69.
[2] G. Hildreth. Learning the Three R's. Minneapolis: Educational Publishers, Inc. 1936, p. 142–143.
[3] A. I. Gates. The Necessary Mental Age for Beginning Reading. Elem. Sch. J. 1937, 37, 497–508.
[4] E. B. Bigelow. School Progress of Underage Children. Elem. Sch. J. 1934, 35, 186–192.
[5] W. W. Wright. Reading Readiness—A Prognostic Study. Bull. Sch. Educ. Ind. Univ., 1936, 12, no. 3, pp. 46.
[6] M. Monroe. Reading Aptitude Tests, New York, Houghton Mifflin Co. 1935.
[7] A. Grant. The Comparative Validity of the Metropolitan Readiness Tests and the Pintner Cunningham Primary Mental Tests. Elem. Sch. J. 1938, 38, 599–605.
[8] G. Hildreth, op. cit., p. 135.

dren with known difficulties as revealed by the readiness examinations and who early exhibited reading difficulties received out-of-classroom tutoring directed toward their difficulties. Furthermore, prior to their placement in first grade, the children had been subjected while in kindergarten to a program of preparation for reading which included the activities usually considered contributing to the development of reading readiness.

The population studied included 109 children admitted to the first grade of the public school of Hastings-on-Hudson. Eighty-nine children had been examined in May and June, 1938 while attending kindergarten, and the remainder, new admissions and several absentees during the spring testing, were examined in September at the opening of school. The examinations administered were the Monroe Reading Aptitude Test[9], and the Kuhlmann–Anderson Intelligence Tests, Grade I–First semester.[10]

From the intelligence tests, the I. Q. and M. A. scores were obtained, M. A. being projected as of September. The reading aptitude test yielded a composite percentile score based on age at time of testing.*

All the children tested were accepted for placement in first grade. These children ranged in age from 5–1 to 7–1 in September. The psychologist on the basis of the tests recommended the section placement for each child. He attempted to adhere to the following classifications. Section I was designed for children with mental age above 6 years, I. Q.'s above 110, and reading aptitude percentiles above 60. It was not always possible to fulfill all three requirements. For example, a child with M. A. 6–6 and I. Q. 124 and reading aptitude percentile only 57 was placed in this section. In such cases it was necessary to consider the three factors together, permitting the excess of one to compensate for a slight deficiency of a second. Children with mental ages from 5–6 to 6–0, I. Q.'s from 95 to 110, and reading aptitude percentiles from 40 to 60 were recommended for placement in Section 2. Here, too, the difficulty described above was experienced and solved in the same way. Thus a child with M. A. 5–7, I. Q. 98, and reading aptitude percentile 72 was placed in this section. The requirements for Section 3 were mental age below 5–6, I. Q. below 95, and reading aptitude percentile below 40. Similar compromises were made here also and, for example, a child with reading aptitude percentile 27, I. Q. 96, and M. A. 5–10 was recommended for placement in this section. These requirements had been established after an inspection of the range and distribution of these scores, so that fewer children would be placed in the third section and more in the other two sections. These recommendations were submitted to the kindergarten teacher and principal, who, on the basis of their more intimate knowledge of the children gained during the school year made several changes. Thus, of the 39 children actually placed in Section I, 5 had been recommended for Section II, of the 39 children placed in Section 2, 3 had been recommended for Section I, and 2 for Section 3, and of the 31 children placed in Section 3, 5 had been recommended for Section 2.

Reading achievement at the end of the first year was measured by the Gates Primary Reading Tests[11] and the Principal's ratings of oral reading. The Gates Tests are silent reading tests and the three types were administered; Type 1 Word Recognition, Type 2 Sentence Reading, and Type 3 Paragraph Reading. The Principal listened to the oral reading of each child and evaluated the child's performance in terms of A, B, C, D, and F.

Results

It is to be remembered that if placement in these three sections was proper, the reading achievement of Section I, the superior group, should be superior to the achievement of Section 2, the average group, which in turn should be superior to Section 3, the inferior group. On the other hand, in view of the purpose of such placement to bring about higher reading achievement, and in view of the preventive and remedial program simultaneously executed, differences between the Sections would tend to be minimized. In all sections, however, failures would be at a minimum if the entire program was valid. It is also important to keep in mind that other factors operated to interfere to an unknown extent with the plans of this experiment.

[9] M. Monroe, op. cit.

[10] F. Kuhlmann and R. G. Anderson. Kuhlmann–Anderson Intelligence Tests, Minneapolis, Educational Test Bureau, 1933.

* This fact may give those children tested at a later date and not yet old enough to be compared to an older age group a slight advantage over children tested earlier in the program.

[11] A. I. Gates. Gates Primary Reading Tests. New York, Bureau of Publications, Teachers College, Columbia University. 1939.

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR AGE, INTELLIGENCE AND READING APTITUDE

		Section 3	Section 2	Section 1
Age in months	Number of children	31	39	39
	Mean	71	74	74
	Standard Deviation	5.4	2.4	3.5
	Range	61– 85	65– 78	64– 76
Mental Age in months	Number of Children*	30	31	37
	Mean	67	79	85
	Standard Deviation	7.4	6.0	7.8
	Range	52– 86	66– 86	70–102
I. Q.	Number of children*	30	31	37
	Mean	96	106	117
	Standard Deviation	11.4	7.4	11.2
	Range	69–132	93–120	93–144
Reading Aptitude Percentiles	Number of children*	31	31	38
	Mean	31	53	69
	Standard Deviation	9.3	9.3	14.7
	Range	15– 53	27– 72	27– 93

* The number of cases varies and is somewhat less than the total register for each section because of absences on the testing days.

These included such factors as the respective teaching abilities of the three first grade teachers, the smaller size of Section 3, the absences of the children, and the effects of illness.

The differences between the three sections may be gleaned from an inspection of Table 1. All three groups were approximately equal with respect to actual ages with means of 71, 74, and 74 months. The overlapping between the three groups is of considerable extent. The range for the third section is larger because of the presence of three children age 61, 83, 85 months. The difference of 3 months between the means of Sections 3 and 2 is not reliable when compared with its standard error (Table 2). The mean age for the entire group is 73 months. It appears, then, that these children are quite typical of first grade entrants with respect to age and that age alone is probably not a factor in any differences between the sections in reading achievement.

The differences of 12 months between the mean mental ages of Sections 2 and 3 and of 6 months between the mean mental ages of Sections 1 and 2 are statistically reliable as shown in Table 2. Although there is some overlapping in the distributions it is not as great as would appear from the ranges presented in Table 1. For example, in Section 3 there is only 1 case with mental age above

TABLE 2

RELIABILITY OF DIFFERENCES IN AGE, INTELLIGENCE AND READING APTITUDE

		Section 2—Section 3	Section 1—Section 2
Age	Difference	3	0
	Standard Error of the Difference	1.1	.7
	Critical Ratio	2.7*	--
Mental Age	Difference	12	6
	Standard Error of the Difference	1.8	1.7
	Critical Ratio	6.7	3.5
I. Q.	Difference	10	11
	Standard Error of the Difference	2.5	2.2
	Critical Ratio	4.0	5.0
Reading Aptitude Percentile	Difference	22	16
	Standard Error of the Difference	2.5	3.3
	Critical Ratio	8.8	4.8

* Differences are considered statistically reliable when the critical ratio is 3 or more.

81 months and in Section 1 there is only 1 case with mental age below 72 months. The mean mental age for the entire group is 78 months and this would lead to the expectation that, on the average, reading achievement for these children should be quite successful. It is expected, too, that these mental age differences between the sections should tend to produce differences between their reading achievement.

I. Q. differences between the sections are also reliable as indicated in Table 2. Here, too, the overlapping is not at all as large as it would appear from the ranges given in Table 1. For example, in Section 3, there are only 3 children with I. Q. above 104, and these are 108, 114, and 113.* In Section 1, there are only 3 children with I. Q. below 103, and these are 101, 96, and 95. If I. Q. is an index of brightness or learning ability, then there should be differences in the average reading achievement of the three sections. The mean intelligence quotient for all the children is 107 and indicates that their reading achievement might be expected to rise slightly above average.

The differences between the mean reading aptitude percentiles of the three sections are also reliable (Table 2). The ranges of the distribution of percentiles presented in Table 1 again mislead one in judging the amount of overlapping. It is much less than these ranges suggest. For example, in Section 3 there is only one child with a percentile above 44, and that is 53. In Section 2, there is only one child with a percentile below 42, and that is 27. Furthermore, there are only 2 children with percentiles above 65, and they are both 72. In Section 1, there are only 3 children with percentiles below 48, and these are 36, 33, and 27. The differences in the mean percentiles between the sections also suggest that the reading achievement between the sections will differ. The average percentile for all the children is 52.*

As expected, differences between the mean reading scores of the three sections were found. However, it can readily be seen from inspecting the data presented in Table 3, that the differences in the means are not nearly as large as one would anticipate from the differences in intelligence and reading readiness. It is believed that the program of placement, prevention, and treatment operated to reduce the differences between the sections. Furthermore, it should be revealed that four children from Section 3 were sent back to kindergarten shortly after the term started. Presumably these children would have received minimum scores on the reading tests if they had survived the year and taken the tests. There was also one child in this section who outright failed the reading tests and received no scores. If these five children could have received scores and could have been included in the computation of the means, and standard deviations, the differences between this Section and the others would have been greater.

* This child was placed in this section because she was only 5–1 chronologically and her reading aptitude percentile was 18.

* This is slightly below the true average for these children. There were, in the population studied, 30 children whose ages at time of testing were between 5 and 5½. The published norms for the reading aptitude tests go down only to age 5½ to 6. In computing percentile ranks for this group, the raw scores were compared with the norms for the 5½ to 6 year group. As a result, they may have been penalized to some extent by this unfavorable comparison.

TABLE 3

MEANS AND STANDARD DEVIATIONS FOR THE GATES PRIMARY READING TESTS

		Section 3	Section 2	Section 1
Type I Word Recognition	N	25	28	35
	Mean	28	32	37
	S. D.	7.4	9.2	9.0
	Range	13–43	13–45	13–47
	Grade Mean	2.43	2.60	3.00
Type II Sentence Reading	N	26	30	36
	Mean	30	32	39
	S. D.	9.0	8.8	6.0
	Range	5–43	11–42	17–45
	Grade Mean	2.40	2.50	3.00
Type III Paragraph Meaning	N	25	30	35
	Mean	16	17	21
	S. D.	4.6	4.8	4.1
	Range	7–23	7–25	12–26
	Grade Mean	2.30	2.40	2.90

How reliable these differences in reading achievement are, may be determined from an examination of Table 4. The means for Section 1 are, of course, all reliably higher than those of Section 3.** Two of the three differences between Sections 1 and 2 are statistically significant. None of the differences between Sections 2 and 3 are reliable. All the differences are in the expected direction and indicate the advisability of the placement of the children in the three sections.

The reading achievement of all three sections as measured by the three tests averages considerably higher than grades 1.9, the norm for the tests administered to first grade classes at the close of the school year.* These grade means are also shown in Table 3, and it can be seen that the lowest is grade 2.30 for Section 3 in Type III Paragraph Reading. This achievement on the part of this section is all the more striking when it is recalled that its average mental age at the beginning of the school year was 5–7, its average I. Q. was only 96, and its average reading aptitude percentile was 31.

** These differences are obvious and were not shown.

* This is according to the procedure described in the manual. See 11, p. 31.

It is interesting to note the number of children in each section who scored below normal on each test. In Table 5, the number of children receiving grade scores below 1.9, considered normal at the time of testing, are shown. These small numbers again indicate the superior reading achieved by the children. For the total group, the highest percentage is only 10 in the case of Type II. When average performance on all three tests is computed, only 5 percent of the children earned a mean grade below the norm. If these children are considered as failures, this percentage is considerably below the estimate of twenty to twenty-five percent of first grade failures cited earlier. Even if the four children, who were sent back to kindergarten after a brief trial of a month in first grade are included in the group of failures, the percent of the total is still much less, being only nine percent. It should be mentioned here that only one child was actually failed at the end of the year and required to repeat first grade.*

One other comparison remains, that of the Principal's ratings of the children's oral

* At the beginning of the second grade, one other child was placed in first grade. In this case, the reason was excessive absence during the first year.

TABLE 4

RELIABILITY OF DIFFERENCES IN READING ACHIEVEMENT

		Section 2—Section 3	Section 1—Section 2
Type I Word Recognition	Difference	4	5
	Standard Error of the Difference	2.3	2.3
	Critical Ratio	1.7	2.2
Type II Sentence Reading	Difference	2	7
	Standard Error of the Difference	2.2	1.9
	Critical Ratio	.9	3.7
Type III Paragraph Reading	Difference	1	4
	Standard Error of the Difference	1.3	1.1
	Critical Ratio	.9	3.6

TABLE 5

NUMBER OF CHILDREN SCORING BELOW GRADE 1.9

	Section 3		Section 2		Section 1		Total	
Test	N	% of children tested	N	% of children tested	N	% of children tested	N	%
Type I Word Recognition	2	8	2	7	1	3	5	5
Type II Sentence Reading	4	15	5	17	0	0	9	10
Type III Paragraph Reading	5	20	2	7	0	0	7	7
Average Grade of 3 tests	3	12	2	7	0	0	5	5

TABLE 6

EVALUATION OF ORAL READING

Ratings	Section 3 N	Section 3 %	Section 2 N	Section 2 %	Section 1 N	Section 1 %
A	4	15	10	36	24	80
B	11	42	6	21	5	17
C	8	31	9	32	1	3
D	2	8	3	11	--	--
F	1	4	--	--	--	--
Total	26	100	28	100	30	100

reading. These data are arranged in Table 6. The ratings were made in terms of A, B, C, D, and F. The results are in keeping with those obtained with the silent reading tests. The one child who received the rating of F is the same child who could do nothing in the Gates Primary Reading Tests and who was required to repeat the first year.

An analysis of the basic variables used in placing the children in first grade did not reveal any single significant or outstanding factor which could be held responsible for the below grade achievement in those children whose average for the three reading tests was below grade 1.9, and also for the return to kindergarten after a short trial in first grade of those children who were sent back. This analysis is shown below. Whichever factor may be considered, chronological age, mental age, I. Q. or reading aptitude, other children with similar scores could be cited who were successful in reading. The only exception, of course, was the child with I. Q. 69.

Child	C. A. in September	M. A. in September	I. Q.	Reading Aptitude Percentile
1—Reading Achievement 1.79	6–1	6–10	112	72
2—Reading Achievement 1.70	6–3	6–1	97	45
3—Reading Achievement 1.80	5–7	5–5	98	40
4—Reading Achievement 1.68	6–1	5–10	96	27
5—Reading Achievement F	6–3	4–4	69	30
6—Returned to Kindergarten	5–7	5–7	100	21
7—Returned to Kindergarten	5–8	5–3	92	27
8—Returned to Kindergarten	5–5	5–3	95	17
9—Returned to Kindergarten	5–11	4–11	83	18

SUMMARY

Children were placed in first grade according to the results of the Kuhlmann–Anderson Intelligence Tests and the Monroe Reading Aptitude Tests. Three sections were established, which could be viewed as superior, average, and inferior groups. In addition, the children had been given a program of reading readiness activities while in kindergarten, and had received preventive and remedial instruction in individual or small groups during the first year whenever it seemed necessary.

The Gates Primary Reading Tests administered at the end of the first year revealed a strong tendency for the Sections to differ in reading achievement in agreement with the establishment of these sections on the basis of mental age, I. Q., and reading aptitude. Ratings of oral reading made by the Principal also indicated that the section with greater intelligence and reading readiness contained the superior readers and the section with lesser intelligence and reading readiness contained the inferior readers.

The reading achievement of the children as a whole, was above the norm for the end of the first grade and is taken as evidence of the success of the entire program including differential placement, reading readiness activity in kindergarten, and remedial and preventive instruction in addition to regular classroom teaching. Although it is advisable to place children in first grade who are mentally older (M. A. above 6), who possess higher I. Q.'s (I. Q. above 100), and greater readiness for reading instruction (reading aptitude percentile above 50), the results support the belief that under a program as described in this report mentally younger children (M. A. below 6), children with lower I. Q.'s (I. Q. below 100), and children somewhat immature for reading (reading aptitude percentile below 50), can successfully be taught to read in the first grade.

THE EFFECTIVENESS OF ORTHOPTIC TRAINING AS A MEANS OF REMEDIAL INSTRUCTION OF READING

SARAH V. APPERSON*

For many years educators have been conducting investigations upon the immensely complex subject of reading.

Up to date these investigations have followed two avenues of approach: the *educational approach* which is the older; and the more recent one, the *eye-movement approach.* In the first the scientist has measured the reading ability of the child by standardized tests and correlated this measure with other factors considered essential to reading. The second approach has accumulated a vast store of information concerning the mechanical performance of the eyes while reading, such as fixations, regressions, eye span and return sweep.

The research here reported was conducted at the University of Nebraska in the year 1937–1938 and the spring of 1939.[1] This research was set up to ascertain if certain types of orthoptic training would affect reading ability. The problem was: would such training bring about more comfort, accuracy and better comprehension in reading? It has been thought that one important factor underlying reading disability might be lack of binocular synchronization, therefore, the idea was to obtain efficient motor habits by training to produce binocular single vision. Actual reading was at no time introduced into the training. Reading results were judged by means of standardized reading tests and ophthalmograph records.

It is assumed that reading is not a single process but one composed of a group of factors which involve the whole individual. This includes the visual mechanism; the central nervous system; the ability to interpret, which depends upon and varies with mental maturity, experience and background; emotional maturity; interests and motives; health and physical development. Factors outside the individual, but having influence upon reading ability are: type, light, material upon which matter is printed and so on. Just how far these factors limit reading or how interdependent they are within and without the individual is not included in this research.

[1] This research was under the direction of Dr. D. A. Worcester of the Department of Educational Psychology, University of Nebraska. It was made possible by a Fellowship grant from the American Research Council of Optometry. The complete study is on file in the office of the Department of Educational Psychology and Measurements at the University of Nebraska.

SELECTION OF SUBJECTS

The three sections of students in this experiment were all enrolled in Teachers College of the University of Nebraska. Section A was taken from the Freshman Orientation course in which there is given special emphasis upon reading and methods of study. Thirty-three of those most in need of remedial work were chosen for the experimental Group. Section B were sophomores enrolled in an introductory course in Educational Psychology. These students were about equally divided among the poor, average, and good readers. Group C were juniors and seniors. Both groups B and C volunteered for the experiment.

Ophthalmograph records were made for all subjects. These records included, words read per minute, fixations, regressions, and comprehension. The comprehension score was derived from the ten true-false questions accompanying the ophthalmograph reading cards.

Each experimental student was matched with a control student from the same class. The criteria used in this matching were: *The Ohio State University Psychological Test, Form 18; Iowa Silent Reading Test, (Advanced), Form A; Chapman–Cook Speed of Reading Test, Form A; or the Minnesota Speed of Reading Test, Form A.*

Table I shows the mean scores of each group on each of the tests used for matching.

THE TESTING INSTRUMENTS

Reading tests used in this experiment were: *Iowa Silent Reading Test, (Advanced), Forms A and B* designed to survey skills essential to effective silent reading; *Chapman–Cook Speed of Reading Test, Forms A and B,* and *Minnesota Speed of Reading Test, Forms A and B,* designed to measure speed, comprehension and accuracy; *Pressey Gen-*

TABLE I

COMPARISON OF MEAN TEST SCORES FOR THE VARIOUS EXPERIMENTAL AND CONTROL GROUPS ENTERING TRAINING

TEST	SECTION A		SECTION B		SECTION C	
	Experimental	Control	Experimental	Control	Experimental	Control
Ohio Psychological Form 18	62.7	65.2	87.6	88.3	69.9	68.2
Iowa Silent Reading, Form A	115.77	117.18	143.34	151.10	133.30	139.50
Minnesota Speed of Reading, Form A	----	----	21.7	22.3	----	----
Chapman-Cook Speed of Reading, Form A	14.8	15.6	----	----	16.8	17.6

eral Reading Test, a survey test. Scholastic aptitude tests were: *Ohio State University Psychological Test, Form 18; Carnegie Mental Ability Test;* and the *Henman–Nelson Test of Mental Ability for College Students.*

The ophthalmograph, upon which weekly tests were made, is a binocular eye-movement camera using the corneal reflection method for obtaining eye-movement photographic records.

THE TRAINING INSTRUMENTS

Some factors other than visual acuity, which might prevent efficient reading are: defective fixation ability, poor projection sense, narrow zone of recognition, and suppression of vision. The instruments to secure higher efficiency in any of these respects were chosen from among those used by optometrists.[2] The instruments were of three types; one giving rotational exercises, a stereoscope which stimulated binocular single vision, and convergent-divergent synchronizers.

Rotational type exercises were administered by use of the Squint Korrector. This instrument has one twenty-four inch disc, slate colored with radiating lines of yellow, blue and red painted on its surface and a circular, four inch fixation target which may be moved nearer and farther from the center of the disc. It has a bright red glass fixation point in the center of the target. This disc is rotated by a variable reversible motor. During the use of this wheel the student wore prism glasses. Through the use of this instrument giving wide rotational exercises, it was expected to stimulate the peripheral areas of vision, to train for exact fixations, and gross superimposition or fusion.

[2] A complete optometrical examination was given at the beginning and end of each period of training by George A. Parkins, O. D. of Ord, Nebraska who also furnished the orthoptic suggestions. The results of the optometrical examination do not enter into the consideration of the group data.

Stereoscopic training was administered by the Helmholz Stereoscope. This instrument is essentially a pair of plus lenses of sufficient strength to neutralize the distance between the eyes and the fixation targets, combined with a pair of mirrors before each eye, set in such a position as to enable the subject to superimpose or fuse the two targets into one. The machine is equipped with a flash which alternates periods of light and dark. The targets consist of reading cards with control marks of one color for the right eye and another color for the left eye. The subject first fixes his eyes on the red, which requires the exercise of positive accommodation, and then on the blue, which requires the exercise of negative accommodation. This serves to establish reciprocal innervations to the intraocular muscles and at the same time trains for the establishment of binocular single vision and fusion or superimposition, which eliminates suppression areas. The flashing light serves to establish accurate and rapid fixations.

The combination training instrument was the Binocular Synchronizer. This instrument is a stereoscope with two five-inch rotating discs, one visible to each eye. It has adjustable center mirrors to permit the image of the one disc to be superimposed upon the other or the two fused. On the face of each disc colored lines are drawn to stimulate peripheral areas of vision. Suitable slots are cut in the face of the disc in which fixation targets may be placed. The instrument is operated by a variable, reversible electric motor. By use of the Binocular Synchronizer, it was hoped to establish a high order of integrated reciprocal innervations to the extraocular muscles. The fixation targets were placed in such a manner as to require the subject to use both eyes if the complete target was to be seen. This placement helped

TABLE II

A COMPARISON OF INITIAL AND FINAL TEST RESULTS, SECTIONS A, B AND C

Name of Test	Section A							Section B							Section C					
	Experimental (N=33)			Control (N=33)				Experimental (N=20)			Control (N=20)				Experimental (N=26)			Control (N=26)		
	Initial	Final	C. R. I. & F.	Initial	Final	C. R. I. & F.	C. R. E. & C.	Initial	Final	C. R. I. & F.	Initial	Final	C. R. I. & F.	C. R. E. & C.	Initial	Final	C. R. I. & F.	Initial	Final	C. R. I. & F.	C. R. E. & C.
Ohio Psychological																				
Mean Score	62.7			65.2				87.6			88.3				69.92	68.26		68.26			
Iowa Silent Reading																				
Mean Score	115.77	138.06	3.32	117.18	136.33	3.01	.33	147.35	151.10		151.20	164.45			133.30	147.69		139.50	150.80		
S. D.	28.75	25.68		26.89	24.72			26.70	22.32	.99	21.85	24.13	1.82	.51	27.75	25.88	1.92	30.47	34.37	1.26	.26
Chapman-Cook Speed of Reading																				
Mean Score	14.8	15.45		15.66	16.81										16.80	21.26		17.61	18.50		
S. D.	3.98	3.82	.07	4.87	4.18	1.03	.33								4.77	6.48	2.34	2.44	6.08	.69	1.76
Minnesota Speed of Reading																				
Mean Score								21.70	23.30		23.30	23.25								
S. D.								6.09	5.96	.84	6.78	7.96	.40	.21						
Ophthalmograph, Rate																				
Mean Score	271.41	360.42		285.41	318.54			259.05	444.20		282.20	300.05			323.84	391.76		290.61	339.23		
S. D.	58.63	112.85	3.99	78.51	87.78	1.61	1.84	79.15	207.93	3.72	65.93	79.11	.77	3.05	81.16	96.78	2.74	67.53	120.37	1.80	.52
Fixations																				
Mean Score	8.85	7.53		8.07	7.65			9.30	6.42		8.38	7.93			6.93	6.52		8.62	7.00		
S. D.	2.07	2.26	2.29	2.38	2.28	.83	1.16	1.94	2.19	4.51	2.19	2.08	.66	2.62	1.25	1.42	1.13	1.72	1.91	3.01	1.75
Regressions																				
Mean Score	1.41	1.14		1.32	1.09			1.55	.88		1.33	1.17			1.30	.95		1.38	1.08		
S. D.	1.03	.76	1.22	1.61	.72	.75	.11	.85	.68	2.79	.72	.90	.62	1.46	.72	.60	1.94	.91	.88	1.22	.16
Comprehension																				
Mean Score	33.33	63.03		4.63	75.15			48.50	81.00		49.00	76.50			71.15	81.53		45.00	80.34		
S. D.	25.90	25.61	4.78	30.14	12.82	5.52	.22	27.79	10.91	4.86	26.81	14.93	6.01	.52	17.41	14.38	2.34	28.32	15.79	5.56	3.22

to eliminate any suppressions and trained the eyes for the binocular single vision. Through the stimulation of the peripheral, central, and para-macular areas it was expected to sensitize the rods and cones and broaden the zone of recognition.

Routine Training Procedure

Routine training procedure consisted of fifteen minute exercises upon each instrument. Each subject received training upon all of the instruments during each period, rotating the order of the use of the instruments daily. Section A received training daily five days per week for forty-five minutes for six weeks. Sections B and C received training forty-five minutes, three times weekly for six weeks. Section A, control group received training in reading in the regular laboratory during the six weeks period. This consisted of vocabulary training, reading to select the central thought, outlining, paragraphing, and an attempt to build up reading background. *A Manual of Reading Exercises for Freshmen* by L. C. Pressey was used. Control Groups of Sections B and C had no training in reading except that incidental to regular college work.

Group Trends

Critical ratios were computed between the means of the first and the final tests for experimental and control groups and between the differences of the experimental and control groups in each section.

A comparison of the means and final score for each of the groups is shown in Table II.

The data recorded in Table II seem to point to the following conclusions.

1. Orthoptic training, without formal educative training, resulted in gains in speed and comprehension in reading, as measured by the tests employed in this research. The gains made by the orthoptically trained group appear to be approximately equal to, or greater than, the gains made by the control groups. The students in the control groups were in Education 30, and therefore, were in general highly motivated. There is evidence to indicate motivation results in some improvement.

2. Orthoptic training gains by volunteer groups were greater and more significant than those made by students required to take the training.

3. Intelligence appeared to be a factor in improved reading ability. Students of high scholastic aptitude, as a rule, improved most.

A Follow-Up Study

A year after the completion of the above investigation, a follow-up study was made of those students still available. In all, twenty-nine of the experimental students responded and came to the laboratory in May 1939 for retesting.

The groups were first given the *Iowa Silent Reading Test,* (Advanced), *Form B* and the *Chapman–Cook Speed of Reading Test, Form B.* These are the same forms that were administered when these students went off the training period the previous year. Ophthalmograph records were also made for each individual of the groups. Each individual's record was compared with the initial and final tests of the previous training period. From the individual records the experiment and control group records were derived. They were, for the most part, the better students of the group. This might, in part, account for their continued improvement which is superior to some of those of the other groups.

Table III shows the gains and losses in rate of reading and in comprehension made by the members of the experimental and control groups who came back a year later for retesting.

Reviewing these data, the following conclusions are reached:

1. A year after orthoptic training, gains in comprehension continued for the Experimental group. However, the Control group, without orthoptic training, showed a similar gain.

2. A year after orthoptic training, gains in rate of reading for the Experimental Group were superior to initial scores. This gain was equalled by the Control Group.

3. A year after orthoptic training, improvements in eye movements, as shown by number of fixations and regressions, was partially lost but still superior to the initial test for the experimental group. This improvement was less for the number of fixations but more in number of regressions for the control group.

TABLE III

GAINS AND LOSSES IN RATE AND COMPREHENSION FOR EXPERIMENTAL AND CONTROL GROUPS A YEAR AFTER ORTHOPTIC TRAINING (MEAN SCORES)

Test	Experimental Group 1937	1938	1939	C. R. I. & F.	Control Group 1937	1938	1939	C. R. I. & F.	C. R. E. & C.
Iowa Silent Reading Mean Score	131.32	142.34	151.03	1.23	117.13	133.30	147.69	1.92	1.22
Chapman-Cook Speed of Reading Mean Score	19.06	18.37	19.95	1.09	15.66	16.80	21.26	2.84	1.18
Ophthalmograph Record Rate Mean Score	292.03	392.76	323.84	2.74	269.05	416.72	361.93	1.40	1.24
Fixations Mean Score	8.25	6.03	7.08	2.84	9.30	8.37	7.93	.66	.90
Regressions Mean Score	1.16	.83	.96	.76	1.31	1.09	1.32	1.22	.55
Comprehension Mean Score	76.50	75.52	75.52	.00	75.45	78.84	80.76	.42	.42
N. Experimental	29								
Control	20								

4. Any attention focused upon reading appeared to result in an improvement in both reading rate and comprehension which gain tended to be partially maintained a year later.

INDIVIDUAL CASE STUDIES

In grouping the data for the study of group trends the individual differences appeared to be concealed. Often the standard deviation was reduced, while the means of the group remained unchanged. This would seem to indicate that some differences were lost in the massing of these data. A study was made therefore of ten cases who manifested deviations of reading behavior. It was desired to see if the improvements or losses made were permanent, and whether a self-interest was present which had stimulated the student to work for improvement. The following table shows the initial test scores, resultant scores after training, the follow-up test score and the gain or loss between initial and follow-up scores of the ten cases studied.

A brief survey of two of the most interesting case studies follows:

Case A made the greatest gain in rate of reading of any one while in training. Her rate jumped from 428 words per minute to 1090 words as measured by the ophthalmograph. Although she lost this remarkable gain in reading rate a year later, she still retained the considerable advancement of 278 words per minute over her initial rate.

According to the *Ohio State University Psychological Test* and the *Carnegie Mental Ability Test*, this student was of very superior scholastic aptitude; however, her school work did not reflect her high test scores. Her grades ranged from sixty-six to ninety-seven with a mean score of eighty.

A's initial score on the *Iowa Silent Reading Test, (Advanced), Form A* was 182. At the end of training, she raised her score on *Form B* to 187. A year later she made a score of 198 on the same *Form B* test. Her superiority in reading was confirmed by *The Pressey General Reading Test*, which showed Student A to be in the ninety-fifth percentile. Her grades on the *Chapman–Cook Speed of Reading Test* were twenty-six, twenty-eight and a year later thirty or a perfect score.

This student was rather timid with strangers, particularly adults, seldom volunteering to take the initiative in class, a fact which may have contributed to her low grades. Upon her entrance into the University, *A* lived in an organized, honor house with fifteen other girls where she became a leader in house activities. She was not however, prominent in University extra-curricular activities, for she lacked the necessary initiative and self-confidence. She preferred books to people and read constantly.

Student A showed good home adjustment, good health adjustment, but a definite timidity with respect to the future and her ability to cope with it. She was very co-operative at

TABLE IV

RESULTS OF TESTS GIVEN TO THE TEN INDIVIDUAL CASES

Case	Iowa Silent Reading			Chapman-Cook			Ophthalmograph Rate			Fixations			Regressions			Comprehension			Ohio Psychological
	1937	1938	1939	Diff.	1937	1938	1939	Diff.	1937	1938	1939	Diff.	1937	1938	1939	Diff.	1937	1938	Diff.	1937
A	182	187	198	+16	26	28	30	+4	428	1090	760	+332	5.8	3.2	5.0	+1.8	4.	0.	4.	80	80	80	0	101
B	150	156	162	+12	27	28	30	+3	286	570	400	+114	7.6	3.6	6.0	+1.6	2.	2.	.6	80	100	70	−10	105
C	130	147	151	+21	15	21	23	+8	171	587	330	+159	7.4	6.6	7.1	+.3	1.6	.6	1.5	80	90	90	+10	82
D	96	127	95	−1	15	12	15	0	261	516	292	+31	10.2	6.8	9.0	+1.2	2.	1.	1.4	20	80	80	+60	32
E	145	166	107	−38	13	18	14	+1	257	461	257	0	9.0	6.0	8.0	−1.	1.6	.6	.8	0	70	40	+40	53
F	145	158	163	+18	28	21	30	+2	160	545	395	+235	12.6	5.2	8.6	+3.4	3.	.4	2.6	100	70	90	−10	82
G	112	134	166	+56	26	21	20	−6	200	200	522	+322	10.0	5.0	4.4	−5.6	2.	.4	2.	20	90	70	+50	107
H	170	152	170	0	15	25	28	+13	340	340	334	−6	6.4	6.6	7.4	+1.	2.2	1.	.4	60	100	90	+40	62
I	121	158	161	+40	21	14	19	−2	387	216	385	−2	8.2	7.0	8.0	−.2	1.2	1.4	.2	90	70	90	0	96
K	102	110	112	+10	14	14	20	+6	464	248	276	−191	5.6	5.6	8.4	+2.8	1.	.6	.6	90	70	50	−40	57

all times and most anxious to succeed. A little praise spurred her to greater efforts and success brought forth more success.

A is typical of those cases whose scholastic aptitudes are far superior to their achievement in reading, but who responded to training by a large increase in rate of reading.

Case K, was older than most of the students, he was twenty-six. According to the *Bell Adjustment Inventory*, he was well adapted at home and in good health. His social adaption was average, but his emotional adjustment was unsatisfactory. His social adaptation did not seem "average", for he did not mingle with any of the students, was inclined to be worried and depressed. This was no doubt due in part to the fact that he carried a full school load while working in a bakery until three in the morning.

The *Ohio State University Psychological Test* and the *Carnegie Test of Mental Ability* showed him to be of low scholastic aptitude.

His attitude toward the experiment was one of willingness, but he was decidedly bored and saw no reason for it. He cared nothing at all for reading, read as little as possible to "get by", so why spend time improving his method.

Student K lost consistently during training in rate as measured by the ophthalmograph, going from 464 words on the initial test to 248 words on the final test and making a gain of only twenty-eight words over the initial test the following year. His comprehension dropped from ninety to seventy during training and a year later dropped still lower to fifty, as measured by the ophthalmograph cards.

However, his *Iowa Silent Reading Test* showed a slight increase in comprehension going from 102 on the initial test to 110 on the final and 112 on the follow-up test. Student K showed a similar small gain in the *Chapman–Cook Speed of Reading Test*, his initial and final scores on the tests taken during training were fourteen, but his follow-up score was twenty points.

Student K quit school the following year to be married and is now a professional cook. He is still wondering why he wasted his time in school when he might be head cook now. He is a typical example of the student who goes to school without motive or definite plan.

The complete data of these ten special cases, lead to these conclusions:

1. Improvement in reading rate and comprehension depends upon many contributing factors differing in individual cases. Motivation and emotional adjustment appear to play an important part.

2. Gains in reading rate and comprehension made during training were not fully maintained a year later. However, in seven cases out of the ten, the increases were retained in part.

3. Students with high scholastic aptitude showed a greater gain than those of low scholastic aptitude.

4. A self-interest which continues to work for improvement appeared to be present in only three cases.

THE INFLUENCE OF ORTHOPTIC TRAINING ON THE READING ABILITY OF COLLEGE FRESHMEN

D. A. WORCESTER
University of Nebraska[1]

The research here reported is really supplementary to that of Sarah V. Apperson, *The Effectiveness of Orthoptic Training As a Means of Remedial Instruction of Reading*, and the orthoptic instruments used were those which are described in her paper, i. e. (1) *The Squint Korrector*, a rotational-type instrument, the purpose of which is to broaden the zone of recognition, to build integrated reciprocal innervation to the extra-ocular muscles of the eyes both monocularly and binocularly. (2) *The Binocular Synchronizer*, a combination of a rotation-type and stereoscopic-type instrument, the purpose of which is like that of The Squint Korrector, to broaden the zone of recognition, condition integrated reciprocal innervation to the extra-ocular muscles of the eyes monocularly and binocularly, stimulate paramacular fusion, and to stimulate stereopsis. (3) The Stereoscope (Helmholtz mirror-type), designed to break down central suppressions, decrease reaction time, and balance the intraocular innervation to the ciliary muscle.

SELECTION OF SUBJECT

All students participating in this investigation were freshmen students registered in Education 30, a course in the introduction to teaching in Teachers College at the University of Nebraska with the exception of one group used as an absolute control group. This group was selected at random from Psychology 90, a sophomore and junior course in experimental psychology in the College of Arts and Science.

EXPERIMENT A

From the 273 students registered in Education 30, ninety-nine were selected on the basis of relatively low scores on the Chapman–Cook Speed of Reading test, the Ohio State Psychological examination, Pressey's Reading test, and ophthalmographic records. The scores of these subjects as compared with those of the class as a whole are shown in columns one and two of Table I.

These ninety-nine students were divided into three groups of equal size on a basis of individual matching on the Ohio Psychological Examination. These groups are hereinafter designated as experimental groups A, B, and C. If there were any discrepancies between the scores made by any three matched individuals, the tendency was to place the lowest subject in Group A. The comparative initial scores of these groups are given in columns 3, 8, and 13 of Table I. As seen in this table, the means and standard deviations of the raw scores of the subjects in Groups A, B, and C are similar, the means being 59.15; 61.54; and 62.33 respectively. The Chapman–Cook Speed of Reading test, the Pressey General Reading test, and the ophthalmograph measures of reading, show that these groups were about equal in the reading abilities measured by those instruments. The only difference to be noted is that the subjects in Group A made slightly, but not notably, lower scores on the average than the other groups.

All students were given an optometric examination by H. B. Peters, (O. D., University of California). The initial tests, mentioned above, by which the groups were selected were repeated with all groups at the end of six weeks and were again repeated at the end of the semester, with the exception that the second experimental examination was not given to Group C. Form B of the Pressey and the Chapman–Cook tests were used on this last occasion. Ophthalmographic records were taken once every week on each individual in Groups A and B and the results of these tests made known to the students. In addition to these, the following procedure was employed:

Group A

a. Training was administered in half-hour sessions three times per week for six weeks,

[1] The data for this research were gathered by H. B. Peters, O. D. under a grant from the American Research Council of Optometry. The complete study is available in the office of D. A. Worcester, University of Nebraska.

making a total of eighteen-training sessions. Each session consisted of a training period on the Squint Korrector, fifteen minutes, and a period on the Binocular Synchronizer, fifteen minutes. No attempt was made to keep the order of these instruments constant but both were employed during each training session for each individual. The procedure and purpose of the training was explained in detail to the students as a motivating device so as to maintain their interest and active cooperation in the routine-type training.

d. These students were then returned to the regular laboratory meetings of Education 30 where they were taught reading skills such as scanning, outlining, use of the library, etc., for a period of four weeks. This training consisted largely of exercises from *A Manual of Reading Exercises for Freshmen* by L. C. Pressey.

Group B

a. Training was given in the regular Education 30 laboratory in one two-hour session each week for a period of six weeks. The procedure consisted, in part, in training the skills of reading, in such techniques as scanning, outlining, paragraph meanings, key words, use of the library, etc.

b. These students were then given visual training in the special laboratory meeting for half an hour three times per week for three weeks.

Group C

a. Training was given in the regular Education 30 laboratory in one two-hour session each week during the entire experiment. The procedure consisted, in part, in training the skills of reading in such techniques as scanning, outlining, paragraph meanings, key words, use of the library, etc.

Many factors may influence the scores made by these individuals on the reading tests other than the orthoptic training; and an attempt was made to control, insofar as possible, some of the major ones. One of these controls established was Group J, a group which pursued regular university courses but was given no formal reading or orthoptic training. Two forms of the various reading tests used in this study were given to the group with a period between administrations corresponding to the experimental training period. The results of this testing program are shown in columns 18, 19, and 20, Table I. It is observed from the treatment of the data that the group registered a slight reduction in score on the Chapman–Cook Speed of Reading test but that this difference is insignificant when the critical ratio, 0.46, is considered. All parts and the total scores on the Pressey General Reading test show a decided difference. The total scores on the two forms show a statistically significant difference favoring Form A. This difference yields a critical ratio of 3.73 or 9,999 chances in 10,000 that the difference is not due to chance factors. Even the part tests show differences that approach statistical significance, and, in one case, Part II Vocabulary, exceeds 3.0. It may be concluded either that this group with no training went backward, declined in reading ability as measured, or that Form B was significantly more difficult than Form A. Since the other test instruments used indicate very little change in reading ability of these students, the more nearly reasonable assumption is that Form B of the Pressey General Reading test is significantly more difficult than Form A, even though the published norms for the test would indicate that the opposite is true. The scores made on the two forms of the Iowa Silent Reading test verify the statement that there is probably little or no change in the reading ability of this group during the time between administrations of the tests. Both the critical ratio of 1.10 for comprehension and the critical ratio 1.41 for the rate of silent reading show a small increase in the mean scores from Form A to Form B, but these are not considered statistically significant. When the scores made on the ophthalmograph are compared, there is found practically no change in the number of fixations and regressions per line; critical ratios are 0.28 and 0.33 respectively. It is to be noted that the same test exercise, Card A, was used in both instances, so it is reasonable to expect a slight increase in comprehension and rate, critical ratio of 1.99 and critical ratio of 1.09 respectively. When the same exercise is used, no statistically significant difference appears. It is interesting to notice that by repeating the same exercise the rate of reading is increased while the number of fixations remains essentially the same, indicating that the perception time has been decreased.

TABLE I

DISTRIBUTIONS OF INITIAL, MIDDLE, AND FINAL SCORES FOR THE TOTAL GROUP AND FOR SUB-GROUPS OF EXPERIMENT A

Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
					GROUP A					GROUP B					GROUP C				GROUP J	
	Total in Education 30	Selected Subjects	1. Initial Tests Sept., 1938	2. Middle Tests Dec., 1938	3. Final Tests Jan., 1939	Critical Ratio M² and M¹	Critical Ratio M³ and M¹	1. Initial Tests Sept., 1938	2. Middle Tests Dec., 1938	3. Final Tests Jan., 1939	Critical Ratio M² and M¹	Critical Ratio M³ and M¹	1. Initial Tests Sept., 1938	2. Middle Tests Dec., 1938	3. Final Tests Jan., 1939	Critical Ratio M² and M¹	Critical Ratio M³ and M¹	Initial Tests Feb., 1939	Final Tests April, 1939	Critical Ratio
1. Number	273	99	33	33	33			33	33	33			33	33	33			63	63	
2. Ohio Psychological																				
mean	74.69	61.55	59.15					61.54					62.33							
std. deviation	24.53	18.48	18.43					19.09					18.74							
range	20–120	25–96	31–89					27–96					25–91							
3. Chapman-Cook Speed of Reading			Form A	Form A	Form B			Form A	Form A	Form B			Form A	Form A	Form B			Form A	Form B	
mean	16.17	13.79	13.76	17.73	14.73	4.50	1.15	14.15	16.60	16.67	4.03	2.55	13.45	16.62	15.63	2.48	2.20	16.42	16.06	0.46
std. deviation	4.40	4.05	3.56	3.61	3.23			4.29	4.06	3.78			4.21	5.88	3.74			4.86	4.01	
range	1–28	1–23	7–25	12–24	9–24			1–23	10–27	8–24			5–23	9–28	8–26			7–30	5–30	
4. Pressey General Reading			Form A	Form A	Form B													Form A	Form B	
Paragraph Meaning																				
mean	137.60	111.59	45.76	49.03	44.33	2.20	.23	44.48	46.42	48.67	.71	2.00	46.13	48.62	49.88	.97	1.56	59.19	65.73	2.18
std. deviation	29.68	23.59	10.48	8.96	9.23			7.97	13.63	9.11			9.65	10.77	9.58			9.90	8.02	
range	61–207	72–182	22–58	30–64	32–60			28–64	14–68	36–66			28–70	24–66	28–68			26–74	36–70	
Vocabulary																				
mean			27.30	28.88	25.88	.75	.73	29.09	30.42	29.52	.55	.23	30.61	30.69	27.03	.03	1.79	44.47	37.16	3.12
std. deviation			8.42	8.80	7.81			8.64	10.97	6.50			8.87	12.82	7.04			14.44	11.80	
range			12–55	17–56	10–44			14–48	12–56	13–42			9–52	9–57	10–38			18–75	11–62	
Outling																				
mean			39.45	47.61	42.00	2.32	.75	39.00	46.18	40.03	2.05	.31	37.26	45.90	39.38	2.36	.62	57.92	51.86	2.09
std. deviation			14.44	14.05	13.15			14.62	13.81	12.06			14.19	15.52	13.02			13.28	11.41	
range			5–64	20–78	15–70			13–76	13–73	8–63			1–68	12–73	11–57			10–77	27–70	
Total																				
mean			109.64	126.36	112.55	2.82	.49	112.36	123.39	118.21	1.75	1.00	112.76	125.21	116.09	1.99	.60	161.63	144.69	3.73
std. deviation			23.96	24.20	24.16			24.71	26.40	22.70			21.58	27.90	22.16			29.04	21.79	
range			61–171	80–189	68–159			72–182	77–188	66–149			74–168	67–191	76–168			88–211	94–192	
5. Ophthalmograph Records																		Card A	Card A	
Fixations																				
mean			9.28	6.25	6.23	9.77	9.53	9.64	6.40	6.34	5.89	6.60	9.21		7.00		5.02	8.7	8.8	0.28
std. deviation			1.47	.98	1.08			2.51	1.92	1.43			1.80		1.64			1.60	2.28	
range			5.6–11.8	4.4–9.4	4.4–9.0			6.6–15.2	4.8–10.4	4.2–11.0			6.4–12.2		4.2–11.0			5.4–12.6	5.6–15.2	
Regressions																				
mean			1.92	.67	.59	6.94	7.39	2.07	1.11	.67	2.74	5.30	1.79		.70		5.45	1.05	1.09	0.33
std. deviation			.99	.35	.37			1.57	1.27	.49			.98		.50			0.68	0.81	
range			.4–3.6	.2–1.4	.2–1.2			0.0–4.2	0.0–5.8	0.0–2.0			.02–3.2		0.0–2.0			0.0–3.2	0.0–2.6	
Comprehension																				
mean	85.45	81.75	79.70	95.76	87.27	2.01	2.37	79.00	84.55	90.61	1.59	3.73	79.70		82.81		.86	77	81.26	1.89
std. deviation	14.32	15.89	15.27	8.18	10.23			15.64	12.08	8.41			16.11		12.56			13.84	11.34	
range	30–100	30–100	50–100	60–100	60–100			40–100	50–100	80–100			30–100		50–100			40–100	40–100	
Rate																				
mean	303.25	262.24	288.91	424.00	428.39	9.08	8.24	262.85	405.21	434.82	6.62	7.67	265.06		388.22		5.32	258	271.03	1.09
std. deviation	85.56	67.21	73.54	74.47	92.65			62.01	108.79	112.79			65.56		113.46			62.71	70.58	
range	112–662	121–450	148–387	286–608	293–616			169–354	226–545	256–750			121–450		214–632			152–500	128–453	
6. Iowa Silent Reading																		Form A	Form B	
Total Comprehension																				
mean																		158.05	163.35	1.10
std. deviation																		27.48	26.72	
range																		78–201	73–206	
Rate																				
mean																		28.7	30.8	1.41
std. deviation																		7.53	8.25	
range																		14–47	14–48	

TABLE II

DISTRIBUTIONS OF INITIAL, MIDDLE, AND FINAL SCORES FOR THE TOTAL GROUP AND FOR SUB-GROUPS OF EXPERIMENT B

Number	1	2	3	4	5	6	7	8	9	10	11
		GROUP E			GROUP F			GROUP G		GROUP H	
	Initial Test Febr., 1939	Middle Test Apr., 1939	Final Test May, 1939	Initial Test Febr., 1939	Middle Test Apr., 1939	Final Test May, 1939	Initial Test Febr., 1939	Middle Test Apr., 1939	Final Test May, 1939	Initial Test Febr., 1939	Final Test May, 1939
1. Number	8	8	8	7	7	7	8	8	8	13	13
2. Chapman-Cook Speed of Reading	Form A	Form A	Form B	Form A	Form A	Form B	Form A	Form A	Form B	Form A	Form B
mean	14.5	16.5	18.0	14.57	18.57	16.42	15.75	18.5	19.25	17.69	18.76
range	7—19	10—23	11—21	10—22	14—30	14—27	10—24	8—28	11—26	8—27	7—28
3. Pressey General Reading	Form A	Form A	Form B	Form A	Form A	Form B	Form A	Form A	Form B	Form A	Form B
Paragraph Meaning											
mean	54.75	60.50	55.5	58.57	66.85	57.11	48.25	53.63	55.00	54.62	50.92
range	46—62	54—70	48—64	42—72	58—72	52—62	34—56	36—65	48—60	40—70	34—62
Vocabulary											
mean	30.00	34.37	30.37	36.00	38.57	35.57	35.75	33.13	30.75	41.92	35.69
range	7—44	20—46	22—36	25—43	30—47	31—40	23—54	23—57	17—45	28—59	20—48
Outlining											
mean	52.00	58.62	53.50	51.00	55.14	57.57	61.25	56.13	45.13	56.00	47.92
range	25—66	48—75	36—74	25—70	43—75	44—71	20—69	30—68	53—59	34—70	27—63
Total											
mean	136.75	153.50	140.62	148.14	160.57	150.85	133.25	142.88	130.88	152.54	135.31
range	80—169	127—187	116—174	108—188	142—194	135—169	99—163	104—183	100—157	118—174	87—166
4. Iowa Silent Reading	Form A	Form A	Form B	Form A	Form A	Form B	Form A	Form A	Form B	Form A	Form B
Total Comprehension											
mean	132.62	143.87	144.87	152.71	168.57	169.71	141.37	164.62	155.62	154.46	169.84
range	74—178	89—203	103—208	115—187	113—213	139—201	120—179	113—182	135—181	85—197	135—203
Rate											
mean	23.75	32.5	37.62	37.14	39.57	43.42	34.62	43.25	40.87	34.23	46.61
range	15—39	19—47	25—47	22—47	21—67	31—62	24—52	34—68	31—59	22—58	30—56
5. Ophthalmograph Records											
Fixations											
mean	9.62	7.60	6.12	8.71	7.00	5.51	9.35	6.62	5.57	8.50	6.11
range	6.8—12.4	5.6—10.0	5.0— 7.4	5.0—11.8	4.6— 8.2	4.6— 6.6	6.8—11.8	4.6— 6.6	4.4— 7.0	5.4—11.6	4.2— 8.2
Regressions											
mean	1.87	0.90	0.62	1.37	0.68	0.28	1.45	0.47	0.27	1.48	0.96
range	0.6—4.2	0.2—1.6	0.0—1.8	0.2—2.6	0.2—1.4	0.0—0.8	0.4—2.4	0.2—1.0	0.0—1.0	0.2—3.2	0.2—1.8
Comprehension											
mean	92.50	92.50	96.25	92.85	91.42	98.57	87.50	91.25	95.0	95.83	86.66
range	80—100	80—100	90—100	80—100	80—100	90—100	60—100	80—100	90—100	90—100	60—100
Rate											
mean	243.87	342.12	447.50	294.71	355.42	486.14	245.25	359.0	468.87	264.25	404.83
range	172—343	198—545	316—590	190—522	279—511	387—753	207—316	255—632	387—571	189—444	248—510

Results Experiment A

It will be remembered that in this experiment Group A had orthoptic training for six weeks and then pedagogical training for four weeks; Group B had pedagogical training for six weeks, then orthoptic training for four weeks, and that the training for Group C was entirely pedagogical. The results of the study are shown in columns 3 through 17, Table I. One of the major trends found in the data presented is the fact that, contrary to certain critics, the comprehension did not decline in the orthoptically trained Group A. In all tests considered, this group made scores approximately equal to or better than the other groups on reading comprehension and speed tests. After a consideration of the results of Group J on the Pressey General Reading test in which Form B was found to be significantly more difficult than Form A, it seems reasonable to conclude that all groups made substantial gains in comprehension as a result of the semester's training. In comparing the scores made by the subjects in Groups A and B in the initial and middle tests where the difference was in the type of training received, Group A showed greater improvement in all measures except vocabulary on the Pressey test; and although not all these differences are statistically significant, some approach it, while others exceed it. The mechanics of reading seem to show an especially marked improvement in the orthoptically trained group (A). It is interesting to notice that Groups A and B, those photographed each week, exceeded Group C in the mechanics of reading. Apparently the taking and explaining to students of weekly ophthalmograph records aid in perfecting the mechanics of reading, though this is not statistically significant. It is also of interest to notice that, although Group A received no formal reading training during the first six weeks, the rate of reading as measured by the Chapman–Cook Speed of Reading test and the ophthalmograph records show a marked increase in the speed of reading with an increase in comprehension as measured by the Pressey General Reading test approaching statistical significance. The apparent drop in speed of reading at the end of the semester over the middle test for Group A may be explained in part by the test form but probably more important is the fact that this group was given comprehension-skills training in the pedagogical laboratory rather than speed exercises. The only difference to be noticed in contrasting the total training of Groups A and B is the small superiority of Group A in visual functions as indicated by the number of fixations. This may be due to the fact that Group A received orthoptic training for six weeks and Group B only four; but, at any rate, the order of training procedures appears to make little difference. It is to be remembered that all members of Groups A, B, and C were members of Education 30, in which they were highly motivated to attempt to improve their reading ability.

Changes in Visual Functions

While detailed evidence will not be presented in this paper, there are data to show clearly that orthoptic training by itself is accompanied by greater gains in fixations and regressions than the gains made by the group having pedagogical training alone and that the increase in comprehension and rate of those orthoptically trained was approximately equal to that of the pedagogically trained group. These gains were statistically significant for both groups and the order of training, pedagogical and orthoptic, where both were used, was apparently of little consequence. The orthoptic examination given to Group A at the close of the experiment indicated a consistent trend toward better visual functioning. This trend is more noticeable in those subjects possessing the greatest limitations at the beginning of the experiment. The individual employed in Experiment A were definitely in the lower range of scholastic aptitude and reading ability.

Two of the instruments used, The Squint Korrector and the Binocular Synchronizer, were intended to produce similar results and both of these instruments were of the rotational type. Reading changes occurred in all groups considered, but visual changes were made only with Group A. It is obvious, therefore, that many questions were left unanswered. Would students in the higher levels of ability show similar gains? Is one of the instruments used better than the other? Would instruments of this type produce still better results? Would there be measurable visual changes with groups having only

pedagogical training alone? The answers to some of these questions were approached in Experiment B.

Experiment B

Experiment B also employed, as did Experiment A, students in Education 30 at the University of Nebraska. These were students who were entering the Teachers College at the University of Nebraska in the second semester and the numbers in the group were smaller than those in the group of Experiment A. The results, therefore, must be accepted with caution, but at least an approach to the problem is indicated. The groups were divided into three groups E, F, and G. No attempt was made to match the individuals in one group with those of another on the basis of any test scores. Rather, the groups were chosen in such a way to give, if possible, a random sample of the course. Data not presented here indicate that the groups were approximately equal, as groups, on the tests used which were the same as those used in Experiment A. There were eight subjects in each of these groups. Group E was given training on the Squint Korrector and then the Stereoscope. Group F was given training on the Stereoscope and then the Binocular Synchronizer. Group G was given training on the Binocular Synchronizer and then the Stereoscope. The changes in reading ability of these various groups, insofar as measured by the instruments employed, are summarized in Table II. Because of the small number of cases in each group, just the means and ranges are presented. One of the trends indicated by these data is that the orthoptic training groups made greater net gains in scores on the reading tests in almost all instances. Differences in the effectiveness of various orthoptic procedures as manifested by changes in reading ability seem difficult to isolate. There appears, however, to be a difference in the changes in the number of fixations per line. It appears that the rotational type training is accompanied by greater gains than the stereoscopic training. To examine the relative changes in reading ability accompanying the various procedures, group learning curves were plotted from ophthalmographic records taken each week during the experiment. The number of fixations per line for Group G is consistently less than for Group E. Similarly, there appears to be consistently fewer regressions and greater speed of reading in Group G than Group E. A greater gain in comprehension is noticed, though judging this function on a basis of ten true-false questions is not considered very reliable. Since both Group E and G were trained on rotational type instruments and then the stereoscopic instrument, it seems reasonable to conclude that there is a trend indicating that the Binocular Synchronizer is more efficient for this purpose than the Squint Korrector, though no statistical significance could be attached to this difference based on 8 cases in each group. Group F shows a greater variability than either of the other groups and no consistent trends appear.

Visual examinations were administered to all students in Education 30 during the time of this experiment and were repeated for Groups E, F, and G after the first five weeks of training and for all students at the close of the experiment. Those students not included in the experimental group, thirteen in number, served as a control group. The students in this control group had had pedagogical training during the term. It was interesting to note that, although there were no significant variations in most of the measures of visual functioning, the zone of perception was increased, which might be expected from intensive speed drills and so on, and in one case there was developed a complete suppression of vision in one eye. In all of the experimental groups of Experiment B noticeable changes in visual functioning, the direction towards the normal, were observed. It is to be remembered that all the persons reported in this study were members of a group in which definite attempts were made to secure high motivation for the improvement of study habits.

Attention is drawn to the following summary of the results of training:

1. Orthoptic training, not supplemented by formal pedagogical training in reading, is accompanied by gains in both rate and comprehension in reading as measured by the test instruments employed in this study. The gains made by the orthoptically-trained group appear to be approximately equal to or greater than those accompanying pedagogical training alone. There is evidence to indicate

that both types of training result in significant improvement.

2. Gains in reading ability accompanying orthoptic training are found not only with selected poor readers but also with a random selection of subjects. In the latter case, the gains are even more noticeable.
3. The order of training, where both pedagogical and orthoptic procedures are used during a training program, is apparently of little consequence.

From a consideration of the changes in visual functions manifested by the groups investigated on the items measured, the following group trends appear significant:

1. Orthoptic training is accompanied by significant visual changes that tend toward the criteria of normality set-up.
2. The changes show a consistent trend toward the broadening of visual functions and reducing of visual limitations manifested at the start of the experiment.
3. The dispersion within the group was significantly reduced on nearly all visual measures, indicating that the group became more homogenous.
4. There is relatively no change in the visual functions accompanying pedagogical training alone.
5. Certain functions show the influence of specific instrument training.

In the above experiment, as in most similar studies reported, the length of the reading periods during which observations were made were relatively short. The fact that the general visual functions were improved by training suggests that orthoptic training may have particular value in the course of difficulties which appear significantly only after a long period of reading—perhaps of three or four hours of duration. It is hoped that an experiment along this line will be made in the near future.

Individual Training

The results reported above have been in terms of grouped data only. It is obvious that, especially with the small number of cases used in these studies, certain individual results might deviate very widely from group tendencies. Case studies, therefore, were made of many of the individuals used in the experiment. Cases were selected which were typical of the following conditions:

Case A: Large Change in Reading and Visual Functions

Case B: Large Change in Reading with Little Change in Visual Functions

Case C: Little Change in Reading with Large Change in Visual Functions

Case D: Little Change in Reading and Visual Functions

Case E: Toxic Condition

Case F: Anatomical Deficiency

Case G: Exemplifying Changes Evidenced by Group E

Case H: Exemplifying Changes Evidenced by Group F

Case I: Exemplifying Changes Evidenced by Group G

It might be noted in passing that in the course of the experiment, several cases of toxic sinusitis, glandular disfunction and similar conditions affecting the health were discovered and referred for medical examination.

The detailed case histories will not be presented here, but some of the findings from them are stated below.

1. Orthoptic training seems to be more effective with those students whose scholastic aptitude is superior to their reading ability and who manifest certain functional limitations in the visual mechanism.
2. Those students make the least improvement by the orthoptic training method whose scholastic aptitude is lower than their reading ability and who manifest only slight visual limitations.
3. Subjects of high scholastic aptitude appear to make greater improvement in reading ability and visual functions than subjects of low scholastic aptitude.
4. Motivation appears to be a significant factor in obtaining both reading and visual changes by the orthoptic method. The effect of personality is probably important in relation to this factor, especially in certain cases.

5. Gross anatomical and physical deviations in the visual mechanism may inhibit the development of normal reading ability.
6. In the opinion of the author, the most effective instrument routine in overcoming visual limitations and increasing reading ability of those used in this study is that used with Group G, Experiment B: The Binocular Synchronizer and the Stereoscope.

In consideration of these findings it would seem, then, that the use of orthoptic methods, as indeed the use of any other type of method, should be adapted specifically to the individual needs of the subject.

ANALYSIS OF A GUIDANCE QUESTIONNAIRE FOR STUDENTS OF SPEECH*

HOWARD GILKINSON
University of Minnesota
AND
FRANKLIN H. KNOWER
State University of Iowa

The present study was undertaken with the purpose of developing a guidance questionnaire which could be used to obtain a comprehensive picture of characteristics of the case histories of students of speech. A survey of the literature revealed several questionnaire forms and frequent reference to their use, but little specific data on the development of the form or on the significance of the items selected. Most of these questionnaires were so constructed that the student merely inserted his answer to an item in a blank space provided in the manner of the completion test. This resulted in a considerable amount of variability in interpretation, and subjectivity in the interpretation of the items by the person using the blank for guidance purposes. This situation prompted the attempt to formulate a questionnaire of the multiple choice type wherein the answers are merely checked. Such answers can be subjected to statistical treatment with somewhat greater objectivity than answers which must be interpreted for purposes of classification.

The steps taken in the development of the questionnaire were: first the selection of factors in the background and personality of the speech student which bore reasonable presumption that they might provide useful data for diagnosis and interpretation of speech behavior; second, casting these items into a form which could be easily recorded and classified; third, selecting groups of student speakers by the means of outside criteria which indicated with high probability that they could be classified as good speaker and poor speaker criterion groups; fourth, administering the questionnaire; and fifth, determining the normal responses for such groups and computing the significance of the differences between them.

Three forms of the questionnaire reported here have been developed and subjected to such analysis. Form A contained 64 items, Form B contained 54 items, and in Form C 40 items have been retained.[1, 2] The items in this form have been validated on both high school and college speech students. The number of college students subjected to this analysis was 75 men and 75 women who were classified as inferior speakers and 75 men and 75 women who were classified as superior speakers. There were 100 high school students in each corresponding group. In all there were 350 inferior speakers and 350 superior speakers.

The method used in differentiating superior from inferior speech performers in the college group was a rating procedure in which teachers of speech rated classroom performance in speaking on a graphic rating scale. These ratings were supplemented with an accumulative speech performance grade record over the period of one semester. The corrected reliability coefficient for the ratings of the two critics was +77. The students with the lowest average rating and whose grade for the semester was not above a "C" were selected as the inferior speakers. Students with the highest average rating whose grade for the semester was not lower than "B" were selected as the superior speakers. The addition of the accumulative grade record should have somewhat increased the reliability of the rating classifications.

Students in the high school groups were selected by combining speech ratings by their instructors, self-ratings, percentiles on the

* The authors are indebted for assistance in this study to workers supplied by the Works Progress Administration, Project No. 5209, Sub-Project No. 96.

[1] Gilkinson, Howard, and Knower, F. H., Psychological Studies of Individual Differences Among Students of Speech —Part II, University of Minnesota, Department of Speech, 1939, pp. 197.

[2] Gilkinson, Howard, and Knower, F. H., A Guidance Questionnaire for Students of Speech, C. H. Stoelting Company, Chicago, 1940.

Knower Speech Attitude Scale, and the Knower Speech Experience Inventory. Students were classified as superior students only when they had a high rating by the instructor and on the self evaluation scale; when the score on the Speech Attitude Scale indicated that they liked to talk; and when the score on the Speech Experience Inventory indicated that they had had much experience in speaking. Students selected as inferior speakers were so classified when the reversal of the conditions of classification as superior speakers held true.

A general classification of the items retained in the scale is indicated by the following table:

TABLE I

CLASSIFICATION OF ITEMS

Incidental items	6
Family	3
Vocational interests	2
Social interests and habits	3
Speech experiences and attitudes and interests	14
School and intellectual interests	2
Health, energy, and physical activities	3
Emotional stability and morale	7

A manual for the evaluation of all answers to items is published with the questionnaire by C. H. Stoelting and Company. This report is therefore confined to brief comments on the trends of data and items of greatest differentiating power in various item classifications and subject groups. The first six items of the scale are primarily identification and orientation items not subject to critical differentiation. The socio-economic status of the family determined by occupational classification on the Minnesota Occupational Rating Scale differentiates well for high school students, but only to a slight extent for college students. Being an only child apparently has more relationship to good speech than being an eldest child, where the relationship tends toward poor speech. The father's public speaking activities show definite and high relationship to the speech of the student. Student vocational choice when classified on an occupational rating scale suggests that there is some correlation between level of aspiration and speech performance. The socialization of the student as indicated by the number of new acquaintances among students and faculty was found to be significant in most group comparisons. Having leadership experience is closely related to the quality of speech performance. The necessity for repetition of remarks to make one's self understood and inability to carry a tune indicate voice deficiencies which are diagnostic of speech problems.

A number of items dealing with speech experiences, interests, and self-ratings are definitely significant in indicating qualities of speech performance. Good speakers at both high school and college levels have done more public speaking and acting, have spent more time in speech classes, are more interested in speech activities, and rate themselves as having fewer speech problems than do the poor speakers. Good speakers, particularly at the high school level, report more active reading habits than do poor speakers. All groups, except the college women's group, who report themselves as having considerable pep had higher speech ratings than those who reported themselves as lacking in this quality.

Reported condition of the teeth and quality of hearing had some critical differentiating power in most groups. General morale as indicated by optimism, pride in past, and freedom from tendency to be embarrassed was correlated with good speech in most groups. Good speakers were also usually persons who reported favorable moods and freedom from fears and sensitivities. In a check list of common fears, fear of speaking in public was the only one of relatively high frequency which was not associated with common causes of violent death.

In a personality trait check list poor speakers quite regularly checked themselves as shy, socially sensitive, and self-conscious. Good speakers checked themselves as being friendly, persistent, impulsive, observant, and tactful much more frequently than did poor speakers.

In summary, it may be said that an objective type guidance questionnaire for students of speech has been constructed and validated. Significant items in various areas of the study of personality have been analyzed and shown to be related to the quality of speech performance. The questionnaire should be useful in courses in speech where the methods used include personnel techniques, and for research purposes.

THE EFFECT OF DEFINITE VERSUS INDEFINITE QUANTITATIVE TERMS UPON THE COMPREHENSION AND RETENTION OF SOCIAL STUDIES MATERIAL

OTTO J. GABEL
*Northern Illinois State Teachers College**

The purpose of the present investigation was to attack one phase of the general problem of comprehension with special reference to (1) the relative efficiency of definite as against indefinite methods of presentation of quantitative concepts in social studies material; (2) the relation of the ability of pupils to comprehend and to retain quantitative terms to their intelligence as measured by the I. Q.; and (3) an analysis of the various combinations of response made by pupils to a test designed to measure the retention as well as the application of the meaning of quantitative terms in given situations.

THE PROBLEM

Basically, skill in comprehension rests upon skill in interpreting words, the symbols or vehicles of thoughts and ideas. Nearly every word has many meanings and many uses. Hence, unless the reader brings to words on the printed page meanings which at least approximate those which the writer had in mind, comprehension results in relatively few, incorrect, or definitely erroneous ideas.

Among the words that have great range of meaning are those denoting quantity. Social studies material employs many concepts of time, area, distance, and size which must be expressed in quantitative terms. Two types of quantitative terms are commonly employed; namely, definite and indefinite. Such terms as "1852" or "five feet" are considered definite, and such terms as "many," "few," and "long ago" are considered indefinite. Because the two types of quantitative terms have different ranges of possible meanings, it is possible that they offer varying degrees of difficulty to pupils. However, Bedwell,[1] Cummings,[2] Ryan,[4] and others point out that writers of social studies material apparently are not aware of this fact. Careful analysis of their material reveals that they employ far more indefinite quantitative terms than definite quantitative terms. This raises the important question as to what effect the method of presentation of concepts of quantity has upon comprehension and retention.

METHODS OF ATTACK

1. *Selections Used in the Experiment*

The materials employed to test the comprehension of quantitative concepts of children in the study were typical of those with which pupils come into daily contact in their school work. These materials were written by the present writer and consisted of six one-page selections on subjects representative of the units usually taught in the social studies. The selections contained forty quantitative terms dealing with time, area, distance, and size. Two forms of each selection were prepared, identical word for word except for the fact that in one form the quantitative terms were stated definitely, in the other form indefinitely.

The two sample paragraphs below illustrate how the quantitative terms were employed definitely and indefinitely, respectively, in the selections described above.

The community forest idea is not new. The American colonists brought it with them from Europe. For this reason, there are some community forests in New England which have been in existence *since 1710*. Today the United States Forest Service reports that *more than 1500* forests have been started in at least 27 states, principally in New York, New England, and the lake states.

The community forest idea is not new. The American colonists brought it with them from Europe. For this reason, there are some community forests in New England which have been in existence *since colonial days*. Today the United States Forest Service reports that *a large number* of forests have been started in at least 27 states, principally in New York, New England, and the lake states.

* This article reports, in part, an investigation conducted by the writer as a partial fulfillment of the requirements for the Ph. D. degree in the College of Education of the State University of Iowa. The assistance of Dr. Ernest Horn, under whose direction the study was conducted, is gratefully acknowledged.

2. Construction of the Tests

Two forms of a multiple response test were used to test the comprehension and retention of the quantitative terms used in the selections. Both forms of the test were alike in that the same questions were used in each form. The chief difference was that the foils in one form of the test were stated definitely, and in the other form indefinitely. There was no one to one correspondence of the foils in the two forms of the test except that in the *correct* answers the terms employed were identical with those used in the original selection. The following examples will illustrate.

Type of question employed in the definite test:

About when were community forests first established in New England?
( ) About 1492
(x) About 1710 (Stated in original selection)
( ) About 1800
( ) About 1860

Corresponding question employed in the indefinite test:

About when were community forests first established in New England?
( ) When Columbus discovered America
( ) Shortly after the Revolution
(x) In colonial days (Stated in original selection)
( ) Just before the end of the 19th century

Sets of material were made up in four forms, W, X, Y, and Z. Six reading selections comprised the materials for each form. Each selection dealt with one of the following topics: oil, aviation, forest conservation, coal, safety first, and reaping tools.

For each of these sets of materials the following testing procedure was used. The pupil read the six selections. After reading each selection he immediately took the tests on that selection. The tests covered two separate pages. One page contained the test questions with the definite foils; the other page contained the test questions with the indefinite foils. After answering both pages of test questions on this selection, the pupil immediately turned to the next selection and did as he had done with the first. He continued in this manner to the end of the complete set of materials. No time limit was placed upon reading the selections and taking the tests.

In *Form W* the quantitative terms were presented to the pupils in definite form, then tested first definitely and then indefinitely. In *Form X* the quantitative terms were presented in definite form, then tested first indefinitely and then definitely. In *Form Y* the presentation was in indefinite form and the testing in definite-indefinite order. In *Form Z* the presentation was in indefinite form and the testing in indefinite-definite order.

The following outline shows the general pattern of organization:

Form	Type of presentation	Testing order
W	Definite	Definite-indefinite
X	Definite	Indefinite-definite
Y	Indefinite	Definite-indefinite
Z	Indefinite	Indefinite-definite

Each pupil tested was required to respond to but one of the above described sets of materials. Hence, no single pupil read both definite and indefinite selections but *all pupils* were required to respond to *both* definite and indefinite test items. This testing procedure forced a recognition or retention of certain terms presented, as well as a "transfer" or application of the meaning of these terms to the corresponding items in the opposite series of tests.

In Forms W and X, for example, the concept, "About 1710," was presented definitely in the selections given the pupils to read. In the definite test that followed, the identical words, "About 1710," appeared as one of the four foils of the questions designed to test this concept. Hence, this test was clearly a retention test. In the indefinite test that followed, the above concept appeared in the words, "In colonial days," as one of the four foils in that test. In responding to this test the pupils were required to apply the original concept to this corresponding test item. Hence, the indefinite test forced a transfer or application of meaning, and has therefore been designated as a transfer test.

In Forms Y and Z the concept, "In colonial days," for example, was presented indefinitely in the selections given pupils to read. Hence, in these forms the indefinite test was the retention test and the definite test was the transfer or application test.

3. Population Used in the Experiment

The tests were administered to pupils in each of the four grades 6, 8, 10, and 12 in nine school systems in northern Illinois. These schools provided a total of 1627 pupils,

or approximately 400 pupils in each of the above four grades.

4. Administration of the Tests

The present writer presented the materials personally to all the pupils in the experiment. In all groups the distribution of the four forms of the test was randomized in the following manner: In any given row of pupils, pupil one was given Form W; pupil two, Form X; pupil three, Form Y; pupil four, Form Z; pupil five, Form W; and so on. Hence, every group, in the aggregate, responded to all four forms of the test.

5. Scoring the Tests

The individual score in the test was the number of test items answered correctly. This score consisted of two partial scores, one of which was the number of items answered correctly in that form of the test in which the foils were stated definitely, and the other the number of items answered correctly in the form of the test in which the foils were stated indefinitely. The *combined score* (sum of definite and indefinite) constituted a measure of the total effectiveness of the method of the presentation used in the test.

Results

1. Statistical Analysis of the Results

The primary purpose of the study was to compare the efficiency of the definite as against the indefinite method of presentation as manifested in the mean total (definite plus indefinite) scores of the groups involved in the experiment. Accordingly, the arithmetic means for all groups were computed from the respective raw scores. These are shown in Table 1. Inspection of these data will reveal that the respective mean scores, regardless of grade or type of quantitative concept, are consistently larger in Forms W and X in which the presentation was definite, than in Forms Y and Z in which the presentation was indefinite.

In making the above comparisons, however, it was recognized that there was a possibility of a difference between the groups compared due to factors other than the method of presentation which might have created systematic differences in the experimental groups. Part of any obtained difference as between group means might, of course, be due to chance factors.

It was desired to obtain an estimate of the error of these obtained differences which would take into consideration this chance factor as well as other factors which might have contributed to the differences. These other uncontrolled variables creating real differences in the mean scores were the pupils, grades, and schools. It should be noted, however, that these variables were randomized by virtue of the distribution of the tests as described above.

Accordingly, comparisons between means were made using the analysis of variance technique as described in Lindquist's "Statistical Analysis in Educational Research."[3] This technique, in brief, is essentially a method of analyzing the results from a series of parallel or duplicated experiments and provides an estimate of error which eliminates the effects of systematic differences, such as school differences, from one to another of these duplicated experiments. It also provides an estimate of the errors due to factors, other than the experimental treatment, which create systematic differences in the experimental groups within each duplicated experiment if these factors have been randomized within each experiment.

In one sense, it might be considered that the data analyzed consisted of two separate but simultaneously conducted experiments. In one experiment the *physical* order of testing was held constant; in the other the *psychological* order of testing was held constant. That is, in Forms W and X the material was presented definitely and the pupils in those groups were tested first definitely then indefinitely in Form W, and in the opposite order in Form X; while in Forms Y and Z the material was presented indefinitely, and the pupils in these groups were tested first definitely then indefinitely in Form Y, and in the opposite order in Form Z. Obviously then, a comparison of Form W with Form Y would compare the methods of presentation with the *physical order of testing* (definite-indefinite) held constant. The same would be true in a comparison of Form X with Form Z with the opposite testing order (indefinite-definite) held constant.

However, it was recognized that the definite-indefinite testing order, physically considered, was not the same as what might be called the retention-transfer order which is more in the nature of a psychological

TABLE 1

TOTAL RAW SCORE MEANS BY GRADES, TEST FORMS AND TYPES OF QUANTITATIVE CONCEPTS

Types of Concepts	Form W: Presentation definite; Testing order: definite-indefinite; Grades 6	8	10	12	T	Form X: Presentation definite; Testing order: indefinite-definite; Grades 6	8	10	12	T
Time	12.2	15.3	16.3	18.2	15.5	11.8	14.6	15.6	17.2	14.8
Area	6.9	7.8	9.2	9.9	8.5	6.1	7.6	8.3	8.9	7.7
Distance	11.8	14.2	15.3	16.5	14.5	11.0	13.9	14.8	16.0	13.9
Size	12.8	16.4	17.8	18.9	16.5	12.6	15.7	16.7	18.2	15.8
All	43.8	53.7	58.6	63.5	54.9	41.5	51.9	55.3	60.2	52.2

Types of Concepts	Form Y: Presentation indefinite; Testing order: definite-indefinite; Grades 6	8	10	12	T	Form Z: Presentation indefinite; Testing order: indefinite-definite; Grades 6	8	10	12	T
Time	10.0	11.5	12.8	13.9	12.0	9.5	13.2	13.0	14.6	12.6
Area	5.7	6.3	7.5	8.1	6.9	5.9	7.1	7.8	8.4	7.3
Distance	9.6	10.7	11.8	12.7	11.2	9.5	11.4	12.2	13.0	11.5
Size	10.2	11.3	12.4	13.5	11.8	10.0	12.4	13.1	14.3	12.5
All	35.5	39.8	44.6	48.2	42.0	35.0	44.1	46.2	50.3	44.0

Note: The respective *possible* total raw scores in each form are as follows:

Time.....22 Area.....14 Distance.....20 Size.....24 All.....80

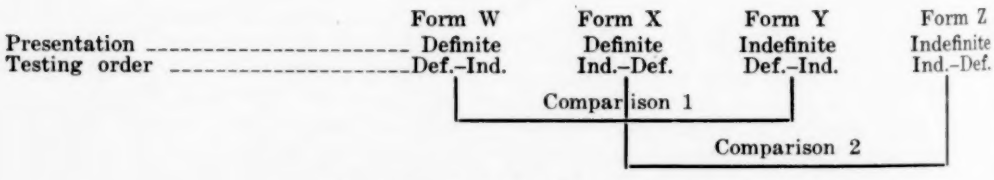
order. By the ***retention*** test is meant that form of the test in which the quantitative terms are the same as those in the materials presented, regardless of whether the test itself is definite or indefinite. Similarly, the ***transfer*** test is that form of the test in which the quantitative terms are expressed in the opposite terminology, (definite, indefinite) to that in the material presented.

Obviously then, regardless of the method of presentation or the physical order of testing, as described above, there is a psychological factor involved in that testing order which subjects the pupil first to the type of the test, be it definite or indefinite, which contains the same terms as those used in the presentation. If the pupil has an opportunity in the first test to retain the material and in the second test to retain it *and* make a transfer to the opposite type of concept, he undergoes a different psychological experience than if he is first asked to retain *and* transfer, and in the second test simply to retain.

Accordingly, Form W (definite presentation) was also compared with Form Z (indefinite presentation) where this psychological order (retention-transfer) was constant. Similarly, Form X was compared with Form Y.

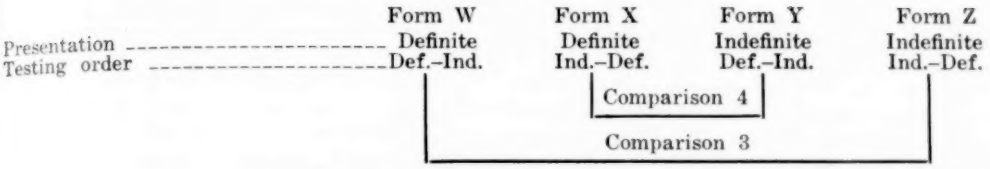
Diagrams A and B will show more clearly the four basic comparisons made, with the two variations of testing order (definite-indefinite), (retention-transfer) held constant in turn.

DIAGRAM A



(Physical order of testing constant)

DIAGRAM B



(Psychological order of testing constant)

Table 2 presents an analysis of the variance of total scores for the four basic comparisons of total scores noted above; namely, Forms W–Y, Forms X–Z, Forms W–Z, and Forms X–Y. As stated above, the analysis of variance technique was used to determine the significance of obtained differences in the comparisons made.

In each of these tables, the difference in mean scores for the two methods of presentation may be regarded as significant at the 5 per cent level if the ratio of the variances for P and P x G x S exceeds 4.26, or at the 1 per cent level if this ratio exceeds 7.82, P having 1 degree of freedom, and P x G x S having 24 degrees of freedom (see Lindquist,

TABLE 2

SUMMARY OF STATISTICAL ANALYSIS OF VARIANCE OF TOTAL SCORES FOR VARIOUS FORMS OF TEST MATERIAL

Forms compared	Sources of variation	Degrees of freedom	Sum of squares	Variance
	P*	1	34,362.10	34,362.10
Total	G	3	29,716.83	9,905.61
	S	8	1,512.04	189.00
Scores	P x G	3	1,431.55	477.18
	P x S	8	492.56	61.57
W—Y	G x S	24	2,021.69	84.24
	P x G x S	24	1,906.84	79.45
	F = P ÷ (P x G x S) = 432.50—(5%—4.26; 1%—7.82)			
	P	1	13,680.55	13,680.55
Total	G	3	31,307.08	10,435.69
	S	8	1,178.60	147.32
Scores	P x G	3	521.33	173.78
	P x S	8	435.03	54.38
X—Z	G x S	24	3,229.67	134.57
	P x G x S	24	1,672.58	69.69
	F = P ÷ (P x G x S) = 196.31—(5%—4.26; 1%—7.82)			
	P	1	24,308.16	24,308.16
Total	G	3	33,240.94	11,080.31
	S	8	1,129.62	141.20
Scores	P x G	3	999.39	333.13
	P x S	8	328.65	41.08
W—Z	G x S	24	3,018.92	125.79
	P x G x S	24	4,794.23	199.76
	F = P ÷ (P x G x S) = 121.69—(5%—4.26; 1%—7.82)			
	P	1	21,306.54	21,306.54
Total	G	3	27,614.00	9,204.67
	S	8	1,520.47	190.06
Scores	P x G	3	1,122.46	374.15
	P x S	8	639.45	79.93
X—Y	G x S	24	1,748.71	72.86
	P x G x S	24	2,167.10	90.30
	F = P ÷ (P x G x S) = 235.95—(5%—4.26; 1%—7.82)			

*Key: P—Presentation
G—Grades
S—Schools
P x G—Presentation by grades
P x S—Presentation by schools
G x S—Grades by schools
P x G x S—Presentation by grades by schools
F—Variance ratio

op. cit., pages 60–65). In each comparison made in Table 2, this variance ratio (F) was many times larger than was necessary to be significant at the 1 per cent level. That is, there are 99 or more chances out of a hundred, or practical certainty, that the obtained differences are not entirely due to chance factors. In other words, there is a significant difference between means in favor of the definite presentation in all comparisons made, i. e., regardless of either of the testing orders (definite-indefinite), (retention-transfer).

It seemed desirable to find whether or not the various types of quantitative concepts offered within themselves any exceptions to the results found in the comparisons of means of the total scores of the test as a whole. The analysis of the data for these various quantitative concepts show that the variance ratio F for the various comparisons made between similar groups as those used for the test as a whole, ranged from 11.96 to 605.02. Hence, without exception, the obtained differences between mean scores made on parts of the test involving various types of quantitative concepts were also significant at the one per cent level. That is to say, the data indicate that the definite method of presentation is superior in effectiveness to the indefinite presentation for concepts involving time, area, distance, or size, whether taken singly or in combination.

For the purpose of presenting the data in a form which is perhaps more comprehensible to the reader, and in a detail which is not feasible with the analysis of variance technique, the percentages of pupils making the correct responses to various forms of test material, by grades, types of quantitative concept, and test form, is presented in Table 3. In this table it will be noted that, for the convenience of the reader, the type of test has been designated as "retention" or "transfer," as the case may be.

Since the superiority of the definite method of presentation has been demonstrated by the analysis of variance technique, using the raw data (total scores, i.e., definite plus indefinite test scores), comment here will be restricted to a discussion of some facts which, in view of the results of previous statistical treatment, it was not considered necessary to deal with in further statistical detail.

It will be noted that the percentages of correct response in the retention tests were higher than those in the transfer tests, regardless of grades, forms of test materials, or types of quantitative concepts. For example, in Table 3 the percentage of correct response for any one of the grades involving any one of the quantitative concepts in the definite (retention) test in Form W is higher than that for any of the percentages in the indefinite (transfer) test in the same form. While the indefinite tests in Form Y and Z show higher percentages than the definite tests, it should be remembered that *in these forms the indefinite test is the retention test.* An examination of the *retention* percentages in Table 3 will reveal a consistent advantage for Forms W and X over the *retention* percentages for Forms Y and Z. The same will be true when the *transfer* percentages are similarly examined. This holds true in favor of the definite method of presentation (Forms W and X) over the indefinite presentation (Forms Y and Z) regardless of grade or type of quantitative concept. The superiority of the definite presentation is not, however, as great in the case of the *transfer* tests as in that of the *retention* tests.

This presentation of the material does not permit a direct comparison of the *total effectiveness* of the methods of presentation, as measured by the ability to respond to *both* types of tests. The previous statistical treatment dealt with this aspect of the situation. It does, however, permit a general inspection of grade differences regardless of the form of test material, the type of quantitative concept, and the order of testing. It is obvious that if material is tested *only* in the manner in which it is presented, the superiority of that method of presentation can be demonstrated. Evidence of this can be found in Table 3 by noting, for example, that the percentages of correct response for any grade or type of quantitative concept in the *definite* test exceed those in the indefinite test if the presentation was *definite;* and the percentages of correct response in the *indefinite* test exceed those in the definite test if the presentation was *indefinite*. Only if two types of presentation are tested in the same *total* way can their relative effectiveness be adequately evaluated. The truth of this statement is attested by the fact that the apparent superiority of the *indefinite* presentation (Forms Y and Z) when tested indefinitely, as shown in Table 3, becomes an inferiority when subjected to the test of total effectiveness.

TABLE 3

PER CENT OF TOTAL NUMBER OF PUPILS ANSWERING THE QUESTIONS WHO GAVE CORRECT RESPONSES IN DEFINITE AND INDEFINITE TESTS, RESPECTIVELY, BY TYPES OF QUANTITATIVE CONCEPTS, TYPES OF TESTS, GRADES, AND TEST FORMS

Form W —Presentation definite
Testing order: Definite (retention)—Indefinite (transfer)

	Time		Area		Distance		Size	
Grade	Def. R*	Ind. T*	Def. R	Ind. T	Def. R	Ind. T	Def. R	Ind. T
6	62.3	48.1	61.0	36.6	65.5	50.7	59.9	46.2
8	78.8	59.8	70.5	40.2	80.7	60.7	78.0	57.7
10	83.8	64.6	81.2	50.2	87.3	65.0	85.3	63.2
12	90.4	75.6	85.7	53.6	94.1	70.6	91.7	65.8
Total	78.8	62.0	75.4	45.2	81.9	61.7	78.7	58.2

Form X —Presentation definite
Testing order: Indefinite (transfer)—Definite (retention)

	Time		Area		Distance		Size	
Grade	Def. R	Ind. T	Def. R	Ind. T	Def. R	Ind. T	Def. R	Ind. T
6	57.7	49.5	48.7	37.8	58.6	50.1	55.3	49.3
8	72.8	59.5	66.4	41.3	75.9	63.5	70.0	60.7
10	77.0	64.3	70.5	47.5	81.6	64.8	78.7	59.7
12	86.0	70.2	74.2	53.3	88.9	70.8	86.0	65.5
Total	73.4	60.9	64.9	45.0	76.3	62.3	72.5	58.8

Form Y —Presentation indefinite
Testing order: Definite (transfer)—Indefinite (retention)

	Time		Area		Distance		Size	
Grade	Def. T	Ind. R	Def. T	Ind. R	Def. T	Ind. R	Def. T	Ind. R
6	32.1	57.6	33.6	47.3	43.6	51.9	35.1	48.9
8	41.2	62.3	39.6	49.1	45.8	61.7	39.5	53.5
10	46.7	69.7	47.0	60.9	51.1	66.4	41.3	62.1
12	52.6	73.9	49.5	65.9	55.7	71.4	46.3	65.6
Total	43.1	65.8	42.4	55.8	49.0	62.8	40.5	57.5

Form Z —Presentation indefinite
Testing order: Indefinite (retention)—Definite (transfer)

	Time		Area		Distance		Size	
Grade	Def. T	Ind. R	Def. T	Ind. R	Def. T	Ind. R	Def. T	Ind. R
6	30.4	55.0	35.3	48.8	41.9	52.6	33.9	49.2
8	48.8	71.0	39.0	61.8	47.6	66.8	42.2	62.1
10	47.7	70.5	45.1	66.8	50.9	71.0	45.3	62.8
12	56.1	76.0	49.8	70.6	54.2	74.9	47.2	70.2
Total	45.8	68.2	42.3	62.1	48.7	66.4	42.2	61.2

* R—Retention; T—Transfer

2. Correlations Between I. Q. and Test Scores

One of the purposes of the study was to determine whether there might not be a relationship between obtained test scores and intelligence. This relationship was not a matter of major concern in the study, but since data on intelligence were available on about 1125 pupils who were used in the study, simple Pearsonian coefficients of correlation were calculated as a possible guide to those who might desire to go into this aspect of the question further.

These correlations were obtained between the I. Q.'s as received from the administrative offices of the schools concerned and the test scores. These I. Q.'s were not derived from the same intelligence tests, but were derived from tests administered under varying conditions and at varying times in the past. Hence, they are probably subject to

more than usual amount of unreliability. Nevertheless, it was felt that these correlations might prove of value as an indication of the necessity of a more complete study.

Correlations were calculated between these I. Q.'s and total scores, i. e., the sum of the definite and indefinite scores, for each of the four forms of test material. Correlations were also calculated between I. Q.'s and the total scores made on the various types of quantitative concepts within each form of test material; namely, time, area, distance, and size. These correlations are shown in Table 4.

All the obtained coefficients of correlation were positive, ranging from .45 ± .03 to .54 ± .03 as between I. Q.'s and total test scores on the respective four forms of the test; and ranging from .43 ± .03 to .50 ± .03; from .29 ± .04 to .37 ± .03; from .35 ± .04 to .43 ± .03; and from .33 ± .04 to .49 ± .03 as between I. Q.'s and total test scores on those parts of the test involving quantitative concepts of time, area, distance, and size, respectively. It will be noted that there is no apparent outstanding difference between the correlation coefficients either as to the total score for the complete forms or for the total scores involving time, area, distance, and size in each of the forms.

The obtained coefficients seem to indicate that there is a strong probability that the more intelligent individuals are likely to make better scores on any form of this test, or on any part of any form, than the less intelligent ones.

TABLE 4

COEFFICIENTS OF CORRELATION, TOGETHER WITH PROBABLE ERRORS, BETWEEN I.Q.'S AND TOTAL SCORES (SUM OF DEFINITE AND INDEFINITE SCORES) BY KIND OF QUANTITATIVE CONCEPT AND FORM OF TEST MATERIAL

Type of test material	Form of test material W	X	Y	Z
Time	.50 ± .03	.45 ± .03	.43 ± .03	.43 ± .03
Area	.36 ± .03	.30 ± .04	.37 ± .03	.29 ± .04
Distance	.43 ± .03	.39 ± .03	.42 ± .03	.35 ± .04
Size	.49 ± .03	.41 ± .03	.33 ± .04	.36 ± .04
All concepts	.52 ± .03	.45 ± .03	.54 ± .03	.45 ± .03

3. Analysis of Response Combinations

A third attempt to treat the data gathered in the present study involved an analysis of the combinations of response made by pupils who took the test. It will be recalled that each pupil was required to respond to each test question definitely as well as indefinitely. Hence, his total response to any given question was a combination of two responses.

These response combinations were recorded in tables (not shown in this brief) which contained a complete record of every response to every question. In other words, these tables show not only the correct response combinations, but also the incorrect response combinations made by the pupils. This was done on the assumption that much may be learned of the nature of the comprehension of quantitative terms through a careful study of the incorrect as well as the correct responses made by pupils to questions upon such terms.

Among the specific analyses made of the recorded response combinations was a study of the correct responses, i. e., the per cent of the total number of pupils answering the questions who gave both definite and indefinite answers correctly. These percentages are shown in Table 5. The data show that the percentages of correct response for all grades, individually or combined, are much larger in Forms W and X in which the presentation was definite than the percentages of correct response in Forms Y and Z in which the presentation was indefinite.

The data also show that the percentages of correct response increase in size from grade to grade in all forms and in all types of quantitative concepts. There is a wider increase, however, between these percentages between the sixth grade and the eighth grade, in all cases, than between the eighth grade and the tenth, or the tenth and the twelfth grades. This may be due to one or more factors; namely, a rather more important change in the type of school experience as between the sixth and eighth grades than as between the eighth and tenth or tenth and twelfth grades; or a natural maturation in the ability to synthesize definite and indefinite meanings

TABLE 5

PER CENT OF TOTAL NUMBER OF PUPILS ANSWERING QUESTIONS WHO GAVE BOTH DEFINITE AND INDEFINITE ANSWERS CORRECTLY TO ALL QUESTIONS, BY GRADES, TEST FORMS, AND TYPES OF QUANTITATIVE CONCEPTS

Form W
Presentation definite
Testing order: definite-indefinite

All Questions	Grades 6	8	10	12	T
Time	36.7	53.2	59.0	70.7	54.9
Area	25.0	32.6	43.5	51.2	38.1
Distance	38.0	52.9	61.4	68.2	55.2
Size	32.1	50.0	57.1	62.8	50.6

Form X
Presentation definite
Testing order: indefinite-definite

	Grades 6	8	10	12	T
Time	34.6	50.3	55.6	66.2	51.7
Area	21.4	33.4	39.6	45.5	35.0
Distance	35.4	53.0	56.1	65.9	52.7
Size	33.5	47.8	51.2	59.6	48.1

Form Y
Presentation indefinite
Testing order: definite-indefinite

All Questions	Grades 6	8	10	12	T
Time	22.9	30.2	37.1	45.4	33.9
Area	20.4	25.9	34.1	39.6	30.0
Distance	26.6	32.0	35.5	40.0	33.5
Size	20.2	25.8	28.1	32.3	26.6

Form Z
Presentation indefinite
Testing order: indefinite-definite

	Grades 6	8	10	12	T
Time	20.0	38.3	38.1	48.2	36.2
Area	21.8	28.0	36.0	41.8	32.0
Distance	24.0	35.4	37.8	42.9	35.1
Size	19.1	29.9	30.6	36.9	29.2

which manifests itself at about the age of the average pupil who is at the seventh grade school level.

The above described data seem to show further detailed evidence that the definite method of presentation is more effective than the indefinite method of presentation of quantitative terms in social studies material.

CONCLUSIONS AND GENERALIZATIONS

The conclusions and generalizations to be drawn from the analysis of the data of the present study are obviously limited to the materials and procedures used. The findings of the study, however, seem to warrant the following:

1. The obtained F (statistical measure of significance) for any of the comparisons made between the mean criterion scores of the various groups in the study was many times as large as was necessary to be significant at the one per cent level. It may be concluded, therefore, that the definite method of presentation of quantitative terms in social studies material is more effective than the indefinite method of presentation. This holds true regardless of the type of quantitative concept or grade level.

2. The obtained coefficients of correlation between obtained test scores and intelligence as measured by the I. Q. were all positive, ranging from .29 to .54 when total scores were used. However, these coefficients were consistent regardless of the type of quantitative concept or type of test material involved. From this fact it may be concluded that there is a strong probability that the more intelligent pupils are more likely to comprehend and retain quantitative concepts in social studies material than the less intelligent ones.

3. The analysis of the response combinations made by the pupils to the tests in the experiment show that the respective percentages of correct response for all grades and for all types of quantitative terms were larger in those forms employing the definite presentation than they were in those forms employing the indefinite presentation. This fact seems to show further evidence of the superiority of the definite presentation over the indefinite method of presentation.

These analyses not only show that the percentages of correct response increase in size from grade to grade in all forms, but also show that these percentages have a wider increase between grades six and eight than between eight and ten, or between ten and twelve. This difference appeared consistently throughout the experiment, and although perhaps not of great importance, it is suggested for further study.

4. The foregoing conclusions suggest definitely that textbook writers should be more discriminating in their selection of terms to express concepts of quantity. In view of the results of the present investigation, a larger number of definite quantitative terms should be employed in social studies material than are being employed at the present time.

Teachers also should recognize the limitations of pupils' comprehension of terms of quantity, and should see that adequate and varied experiential background be provided for understanding quantitative concepts which are essential in reading social studies material.

Bibliography

1. Bedwell, Margaret. *Comprehension of Concepts of Quantity in Third Grade Social Studies Reading Material,* Master's Thesis, University of Iowa, Iowa City, 1932.
2. Cummings, Mable E. *Comprehension as Affected by the Use of Quantitative Terms in Social Science Textbooks,* Master's Thesis, University of Iowa, Iowa City, 1933.
3. Lindquist, E. F. *Statistical Analysis in Educational Research,* Houghton Mifflin Company, Boston, 1940.
4. Ryan, Grace M. *A Study of the Comprehension of Quantitative Terms in Geography at the Fifth Grade Level,* Master's Thesis, University of Iowa, Iowa City, 1935.

THE RELATION OF CERTAIN SOCIAL ATTITUDES TO SCHOOL ACHIEVEMENT

HAROLD JAYNES BINGHAM
George Peabody College for Teachers

The purpose of this study was to consider certain social attitudes as measured by an attitude test in relation to school achievement as measured by grades made in school. As a corollary to the main problem, the relation of information test scores and intelligence test scores to the same social attitudes were considered. Such other items as sex, school progress, and class in school were examined with reference to the general problem of factors related to social attitudes. No attempt was made to determine causal relationships between any factor and the social attitudes considered.

TERMS USED IN THE STUDY

Thurstone's (2,531) definition of "attitude" is used throughout this study. He considers that the term "attitude" denotes: ". . . the sum total of man's inclinations and feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and convictions about any specified topic." These feelings and thoughts of an individual could best be measured through a study of his actions. To test attitudes in such a manner necessitates personal observation. This type of procedure would make investigation difficult. However, by having a subject respond to a list of questions concerning the problems under consideration, his verbally expressed opinions can be reduced to a quantitative estimate for purposes of comparison. A stated opinion is regarded as the verbal expression of an attitude. It was assumed that a high enough relation existed between the verbal expression and the true attitude of the subject for practical use.

"Conservatism" and "liberalism" are used to connote opposite extremes on the scale of social beliefs considered in this study.

"High school" is used to mean the upper three years of a senior high school organized on a six-three-three plan.

"Average grade" refers to the mean of all school marks earned by students since entering senior high school (i. e., seventy, eighty, ninety).

"Intelligence" is used to denote the raw scores made on the Otis Quick Scoring Mental Ability Test, Gamma Form Bm.

"Information" refers to those scores made on an information test concerning the social problems considered in this investigation.

"Class in school" connotes the tenth, eleventh, or twelfth grade.

"School progress" is used to mean the student's rate of progress in school, (i. e., accelerated, average, and retarded).

CONSTRUCTION OF ATTITUDE SCALE

An attitude scale was constructed for each of the following six social problems: war, the foreign policy of the United States, the relation of labor and industry, the government's housing program, federal aid for agriculture, and the government's spending program.

From 150 to 250 statements were secured from articles in leading periodicals on each of these six social problems. After editing these statements, 100 of each group were judged of sufficient pertinency and clarity to be used as a basis from which to build the scales.

Approximately 35 people rated the statements on each social problem on the basis of a scale ranging from one to seven. They were instructed to rate the most conservative statement as one and the most liberal as seven, and to distribute the other statements of varying conservatism-liberalism between the two extremes. The points on the scale were assigned values of one, three, five, seven, nine, eleven, and thirteen. The scale value for each statement was determined by averaging the values assigned by the raters. Each statement was tested for ambiguity by determining the standard deviation of each question. If a statement deviated more than three and one-half points on the scale from its average, it was considered too equivocal for use.

The validity or relevance of the statement was then determined by plotting the statements on a linear scale ranging from the statement with the most conservative scale

value to the statement with the most liberal scale value. The assumption was that relevant statements would fall on or near the linear scale.

The scale for each of the attitudes had different upper and lower limits. For purposes of comparison, all scales were reduced to the same range: from one to eight. Those statements with values of between 3.8 and 5.1 ($\pm$ ½ sigma) on the conservative-liberal scale ranging from one to eight and used in this study are termed "neutral". Those between 5.2 and 6.5 ($+$ ½ sigma to $+$ 1½ sigma) are considered as "liberal", and those 6.6 and above are considered as "more liberal". Those between 3.7 and 2.4 ($-$½ sigma to $-$1½ sigma) are called "conservative", and those with values of 2.3 and below are considered "more conservative". These different degrees of conservatism and liberalism are used merely for describing positions on a sc le of social beliefs. These are relative terms and any statement is held to be "liberal" or "conservative" only in terms of other statements on this particular social attitude scale.

As appearing in the form of a test, there were sixteen statements on war, sixteen on the relation of labor and industry, sixteen on federal aid for agriculture, sixteen, on the government's spending program, seventeen on the foreign policy of the United States, and twenty on the government's housing program. Examples of the statements in the six areas of social attitudes follows:

War:
- War is the greatest handicap to an improving civilization.
- Colonial possessions are necessary to maintain a nation's standard of living.

Relations of Labor and Industry:
- Strikes merely lessen public sympathy for labor.
- Organized labor plans a complete change of our social and economic order.

Federal Aid for Agriculture:
- All farm policies are worked out in the light of election returns.
- The prosperity of the nation is bound up with the prosperity of the farm.

The Government's Housing Program:
- Federal housing projects do not provide suitable homes.
- Improved housing would make the American home more secure.

The Government's Spending Program:
- The financial program of the United States is completely out of control.
- A fundamental change in the ideals of democracy has caused increased government expenditure.

The Foreign Policy of the United States:
- Our country should not fight for any reason short of invasion.
- The armed forces of the United States should be used to police the world.

It is believed that the statements on the attitude test are unambiguous, distribute themselves along the scale from one to eight rather evenly, and are of sufficient validity for the testing of social attitudes.

Procedure

The average grade for each student was determined by computing the mean of all grades made by him since entering high school. The information test scores were secured by administering a self-devised test concerning the six social problems considered in this study. The coefficient of reliability of the test, determined by correlating alternate items, was .80. The intelligence test scores were determined by administering the Otis Quick Scoring Mental Ability Test.

The attitude test, the information test, and the intelligence test were administered to 435 students of Albert G. Parrish high school of Selma, Alabama. There were 356 stude 's for whom all information was complete. It is they who form the basis of this study.

The relation of attitudes toward the six social problems to average grade made in school, intelligence, and information was computed by means of the Pearson Product-Moment method.

On the basis of the grades made in school, each student was classified as A, B, C, D, or E, according to the deviation of his score from the mean of the group. The scores on the information test and also on the intelligence test were the bases for like groupings of the students. The consistency with which each group marked the attitudes of a certain scale value was measured by subtracting the average of the scores of the most conservative statement agreed to by each subject from the average of the scores of the most liberal statement agreed to by each subject.

The relation of each factor to the combined attitudes was determined by computing the arithmetic mean of the separate correlations of the particular factor to each attitude.

Because of the few number of cases the r was corrected according to the following formula in which M is the number of variables.

$$r^2_{yx} = 1 - (1 - r^2_{yx}) \frac{N - 1}{N - M}$$

The relation of the combined factors to each attitude was determined by computing the multiple correlation of these factors to the various attitudes.

The subjects were classified according to sex. Those subjects whose average scale value of a particular social problem was 6.6 points or above were classified as "more liberal", those between 5.2 and 6.5 were classified as "liberal", those between 3.8 and 5.1 as "neutral", those between 2.4 and 3.7 as "conservative", and those 2.3 and below as "more conservative". By the use of the Edgerton–Paterson "Table of Standard Errors and Probable Errors of Percentages for Varying Numbers of Cases", the significance of the differences of the standard error of the percentages was computed for each of the conservative-liberal classifications.

The same statistical procedure was followed in determining the significance of the difference of the attitude scores of Class 10, Class 11, and Class 12, as well as for accelerated, average, and retarded age groups.

Summary of Data

The group as a whole was "neutral" in its attitude toward war, the relation of labor and industry, and the government's spending program; "liberal" in its attitude toward the government's housing program and federal aid for agriculture; and was "conservative" toward the foreign policy of the United States. The average attitude test score for war was 3.84; for the relation of labor and industry, 4.02; for the government's spending program, 5.11; for the government's housing program, 5.56; for federal aid for agriculture, 5.37; and for the foreign policy of the United States, 3.31.

The students with the better grades were in favor of the government's spending program (r, + .32; PE_r, ± .032), and the government's housing program (r, + .31; PE_r, ± .032). They were in favor of industry rather than labor (r, — .19; PE_r, ± .034), and were opposed to an aggressive foreign policy (r, — .25; PE_r, ± .033). There was a tendency, though not definite, for these students to be opposed to war (r, — .13; PE_r, ± .035). An insignificant relationship existed between attitudes toward federal aid for agriculture and grades made in school (r, + .072; PE_r, ± .033).

Those who scored highest on the information test were in favor of the government's housing program (r, + .21; PE_r, ± .033), the government's spending program (r, + .20; PE_r, ± .034), and federal aid for agriculture (r, + .15; PE_r, ± .035). They were opposed to a liberal foreign policy of the United States (r, — .26; PE_r, ± .033). There was a tendency, though not statistically reliable, for those who scored best on the information test to favor industry, (r, — .12; PE_r, ± .035), and to oppose war (r, — .13; PE_r, ± .035).

Those who scored highest on the intelligence test favored the government's housing program (r, + .36; PE_r, ± .031), and were inclined to favor federal aid for agriculture (r, + .11; PE_r, ± .035), and the government's spending program (r, + .12; PE_r, ± .035). They were opposed to war (r, — .18; PE_r, ± .034). The relationship of intelligence test scores to attitudes toward the relation of labor and industry and to the foreign policy of the United States was not significant.

There was a tendency for those who made the highest grades, scored highest on the information test and on the intelligence test to be more consistent in agreement to attitudes of a particular scale value than those who made the lowest grades or scores. For those attitudes where the relation to the various factors was high there was a progressive decline of consistency from group A to group E.

The multiple correlation of grades made in school, information, and intelligence to war was R.18, to the relation of labor and industry was R.20, and to the foreign policy of the United States was R.27. Inasmuch as these multiple correlations were computed from negative zero order correlations, the multiple R's were considered as inverse relations. The combined relation of the factors to the government's spending program was R.42, to the government's housing program

was R.37, and to federal aid for agriculture was R.20. It is noticeable that the relation of these combined factors to the attitudes are little higher than the highest zero order correlations for the same attitude.

When separate zero order correlations of each factor to the various attitudes were combined it was found that grades made in school had a relationship to the combined attitudes of r.21, information a relationship of r.18, and intelligence a relationship of r.15. The correction made for the relation of the combined attitudes to each factor was $_{c}$r.207 for grades and attitudes, $_{c}$r.149 for intelligence and attitudes, and $_{c}$r.175 for information and attitudes. According to Ezekiel's (1,309) interpretation the chances are 60 to 70 in 100 that the corrected relationship of each factor to the combined attitudes is very significant.

Approximately fifty percent of both sexes occupied "neutral" positions in their attitude toward war. The trend was from the "neutral" to a "more conservative" viewpoint. A larger percent of the girls were "conservative" in their viewpoint toward the foreign policy of the United States. Of this sex, 73.3 percent were "conservative", while 68 percent of the boys were "conservative". Of the girls 7.7 percent were "more conservative" and 4.9 percent of the boys were "more conservative" in their attitudes toward the foreign policy of the United States. In their attitudes toward the relation of labor and industry about 66 percent of both sexes were "neutral" in their viewpoint, and about 30 percent of each sex were "conservative". Both the boys and the girls showed favorable attitudes toward federal aid for agriculture; approximately 30 percent were "neutral" in their viewpoint. Eighty-one percent of the girls were "liberal" in their attitude toward the federal housing program, 14 percent more girls than boys were "liberal". There is a wider range of opinion by both sexes in regard to the government's spending program than in the case of the other attitudes showing a positive relationship. Each group ranged from "conservative" to "more liberal".

An examination of the social attitudes of the sexes leaves one more impressed with the similarities of their attitudes than with the differences. By the use of the Edgerton–Paterson table no significant difference was found.

A larger percentage of the students of Classes 11 and 12 were opposed to war than students of Class 10. A progressive increase in the percentage of those "conservative" toward the foreign policy of the United States from Class 10 to Class 12 was apparent. A larger percent of Classes 11 and 12 were "conservative" in their attitude toward the relation of labor and industry than students of Class 10. There was a progressive increase in the percent of those who favored federal aid for agriculture, the government's housing program, and the government's spending program from Class 10 to Class 12. These differences must be regarded merely as tendencies inasmuch as they are not statistically reliable.

A larger percent of the over age group were "neutral" in their attitude than of either of the other two groups except in their attitude toward the government's housing and spending programs. In the former, the same percent of the average group was "neutral", and in the latter, a larger percent of the accelerated group was "neutral". The retarded group was "more liberal" toward the government's spending program.

Conclusions

As long as it is believed that the educational system must cooperate in the shaping of an effective citizenry, the social attitudes held by the products of this system are of importance. To be vital, the school must make clear the workings of democracy in the midst of foreign and domestic crises.

If the opinions of the majority of the American people can be judged by the political and social philosophies which they have supported, this study would tend to substantiate the suggestion that students reveal to a great extent the social attitudes of the society of which they are a part. The liberal viewpoint toward federal aid for agriculture would seem to have typified the special interest of the locale of the study. This liberal tendency and the low relations of agriculture to the factors studied as compared with the relation of these factors to the other social problems might indicate the effect of the long standing cultural interest in agriculture. It is believed that the conservative opinion held toward the foreign policy of the United States reflected the national policy desired at that time.

The low relationship of grades made in school and an information test scores to social attitudes as measured by the attitude test emphasizes the idea that influences other than those measured by quantitative estimates of achievement are the most related to social attitudes.

The writer is inclined to believe that a student may fail to utilize all of the information he possesses in indicating his attitudes on certain social problems. Although the trend was not progressive, the difference in the range of the scores for those with the highest grades and for those with the lowest grades would indicate that the former group is more consistent in its attitudes toward social problems.

Sex is of little importance in determining the social attitudes of secondary school students.

The upper classes are more liberal in their attitudes toward social legislation than in their attitudes toward subjects related to war.

There is a slight tendency for the over age group to occupy neutral positions in their attitudes toward social problems.

One of the functions of education in the United States is to promote intelligent and kindly approaches to the solution of major social problems of this country. Whether or not such solutions are in temporary accordance with what is regarded as either conservative or liberal viewpoints, the low relationships found in this study tend to emphasize the belief that intelligence and information must be made to function more effectively if wise solutions are to be made of the problems of American life.

References

1. Guilford, J. P. *Psychometric Methods*, New York: McGraw–Hill, Inc., 1936, Pp. 309.
2. Thurstone, L. L. "Attitudes can be measured." *Amer. J. Sociol.*, (1928), 33, 529–554.

RELATION BETWEEN SPELLING ABILITY AND READING ABILITY

NELLIE L. PEAKE
Indianapolis Public Schools

The trend in recent word-study books to present new words in context material and to introduce exercises which develop meaning of words has led to a consideration of the relationship between spelling ability and knowledge of word meaning. A testing program which included achievement tests provided opportunity for statistically determining this relationship.

Problem.—This study attempts to answer the following questions:

1. What is the relation between ability in spelling and ability in reading of pupils of grades IV–VIII as measured by scores on the Stanford Achievement Test?
2. Is the correlation relatively uniform throughout the five grades tested?
3. What is the coefficient of correlation for the group as a whole?

Source of material.—All the data for this investigation were obtained directly from the tabulated record of scores made by pupils on the New Stanford Achievement Test, Advanced Form W. This record included scores of every pupil enrolled in grades IV–VIII in an Indianapolis public school. In this way 355 cases were used for investigation. Of these 355 cases, 58 were from Grade IV, 50 from Grade V, 73 from Grade VI, and 87 each from Grades VII and VIII.

Method.—The method employed in this study was statistical since the relationship between the raw scores in spelling and reading was determined by calculating the coefficient of correlation. Coefficients were computed between the two sets of test scores for each half grade, IVB, IVA, etc., and for the entire group as a whole.

Limitations.—The term 'reading ability', as used in this study, refers to only one phase of reading, namely, vocabulary knowledge or word meaning. A further limitation of this investigation is in obtaining spelling and reading scores from only one set of achievement tests.

Definition of terms.—By spelling ability is meant the accuracy with which the sentences in the dictation test were written.

Other similar studies.—That a positive relationship exists between spelling ability

TABLE I

SCORES MADE ON THE STANFORD ACHIEVEMENT SPELLING AND WORD MEANING TESTS BY 355 GRADE SCHOOL PUPILS

	IVB		IVA		VB		VA		VIB		VIA		VIIB		VIIA		VIIIB		VIIIA	
Scores	S	WM	S	WM	S	WM	S	WM	S	WM	S	WM	S	WM	S	WM	S	WM	S	WM
115–119																		1		2
110–114										1		1			1	1	9	2		5
105–109										1			5		2	2	10	2		5
100–104					1		2	1	4		5	1	15		5	8	15	2		9
95– 99		2		1	1		3	3	5	1	4	5	11	2	1	5	9	6	2	
90– 94		7	2		3	1	3	6	10	2	3	7	6	2	4	6	6	3	2	
85– 89		2		1	1	3	1	8	4	5	1	13	6	2	2	13	2	7	2	
80– 84		3		1	2	3		8	5	6		10	6	9	5	12		6	1	
75– 79	5	8	1	3	5	3	2	2	8	9	1	2	8	4	4	4	7	1	3	2
70– 74	5	5		3	4	5	2		6	5	1	1	6	1	2			1		1
65– 69	7	3	6	3	4	6	1	2	6	3	2	3	8	3	3					2
60– 64	8	4	4	1	6	4	1	1		1	1	2	4	2					1	
55– 59	1	1	1	1	6	5			2	1	3			1					1	
50– 54	6		1		1		1		1			2								
45– 49	2	1		4	3	2	1				2	1		2						
40– 44	4	4			1	2	1							1						
35– 39	1	1	4		1															
30– 34	2				1															
Total	41	41	17	17	35	35	15	15	49	49	24	24	63	63	24	24	54	54	33	33

and knowledge of the meaning of the words is noted by Hollingworth[1] as she concludes that ". . . children will produce about sixty-six and two-thirds per cent more of misspellings in writing words of the meaning of which they are ignorant or uncertain, than they will produce in writing words the meaning of which they know."

Tyler,[2] in equating two groups of pupils in carrying out a study of spelling as a secondary learning, found that the correlation between spelling scores and word meaning scores as measured by the Stanford Achievement Test, Form V, was .654 for 151 sixth grade pupils. (Forms V and W of the Stanford Achievement Test have been equated so that scores on Form V have the same value as scores on Form W.)

Findings

The results obtained in this investigation are presented in Tables I and II. Table I shows the distribution of scores for the tests in spelling and word meaning for the 355 cases used in this study. The grades included are from IVB through VIIIA.

Inspection of Table I reveals that the highest spelling scores in each half grade are lower than the highest word meaning scores in every section except one, the VIIB. The exception in this case is due to a single score. (It should be noted here that the individual tests of the Stanford Battery have been equated in such a way that equivalent scores are possible in all the tests.) The range of scores on the two tests shows that the word meaning scores exhibit a wider range of abilities than those of the spelling scores in all except two instances, the VB and the VIIIA. The widest range of abilities is found in the IVB and VIIB word meaning and the VB spelling.

The enrollment by grade sections with the corresponding coefficients of correlation together with the coefficient for the whole group is summarized in Table II.

Table II shows that the highest correlation, .933, was found in IVB, and the lowest, .426, in VIIB. Correlations for the A sections of grades IV, V and VI were about the same, ranging from .811 to .874. The correlations for grades VIIB through VIIIA are lower than any of those found for grades IV–VI.

Table II

Correlations Between Spelling Scores and Word Meaning Scores of 355 Pupils in Grades IV–VIII

Grade	Section	Number	r
IV	B	41	.933
	A	17	.811
V	B	35	.676
	A	15	.830
VI	B	49	.752
	A	24	.874
VII	B	63	.623
	A	24	.426
VIII	B	54	.566
	A	33	.629
Total		355	
Average			.814

The correlation for the whole group was .814 which is fairly high. It indicates that there is considerable agreement between the spelling test scores and the word meaning scores in the group tested.

Conclusions

1. There is a tendency for high scores in spelling to accompany high scores in word meaning. The relatively high positive correlation between test scores in spelling and reading in most of the grades examined is evidence that abilities in these two subjects tend to accompany each other.

2. The coefficients of correlation obtained do not show a uniform amount of correlation throughout the five grades tested since the highest numerical value of r (.933) is over twice as large as the lowest numerical value (.426).

3. The coefficient of correlation for the group as a whole is .814.

[1] L. S. Hollingworth, *The Psychology of Special Disability in Spelling*, Contributions to Education, No. 88. New York: Teachers College, Columbia University, (1918), p. 57.

[2] I. K. Tyler, *Spelling as a Secondary Learning*, Contributions to Education, No. 781. New York: Teachers College, Columbia University, (1939), p. 72.

STATEMENT OF THE OWNERSHIP, MANAGEMENT, CIRCULATION, ETC., REQUIRED BY THE ACTS OF CONGRESS OF AUGUST 24, 1912, AND MARCH 3, 1933

Of Journal of Experimental Education published quarterly at Madison, Wisconsin, for October 1, 1940.

State of Wisconsin—County of Dane—ss.

Before me, a Notary Public in and for the State and county aforesaid, personally appeared A. S. Barr, who, having been duly sworn according to law, deposes and says that he is the owner and editor of the Journal of Experimental Education and that the following is, to the best of his knowledge and belief, a true statement of the ownership, management (and if a daily paper, the circulation), etc., of the aforesaid publication for the date shown in the above caption, required by the Act of August 24, 1912, as amended by the Act of March 3, 1933, embodied in section 537, Postal Laws and Regulations, printed on the reverse of this form, to wit:

1. That the names and addresses of the publisher, editor, managing editor, and business managers are: Publisher, Democrat Printing Co., Madison, Wisconsin; Editor, A. S. Barr, Madison, Wisconsin; Managing Editor, A. S. Barr, Madison, Wisconsin; Business Manager, A. S. Barr, Madison, Wisconsin.

2. That the owner is: (If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding one per cent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a firm, company, or other unincorporated concern, its name and address, as well as those of each individual member, must be given.) A. S. Barr, Dept. of Education, University of Wisconsin, Madison, Wisconsin.

3. That the known bondholders, mortgagees, and other security holders owning or holding 1 per cent or more of total amount of bonds, mortgages, or other securities are: (If there are none, so state.) None.

4. That the two paragraphs next above, giving the names of the owners, stockholders, and security holders, if any, contain not only the list of stockholders and security holders as they appear upon the books of the company but also, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation, the name of the person or corporation for whom such trustee is acting, is given; also that the said two paragraphs contain statements embracing affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner; and this affiant has no reason to believe that any other person, association, or corporation has any interest direct or indirect in the said stock, bonds, or other securities than as so stated by him.

5. That the average number of copies of each issue of this publication sold or distributed, through the mails or otherwise, to paid subscribers during the twelve months preceding the date shown above is..... (This information is required from daily publications only.)

A. S. BARR.

Sworn to and subscribed before me this 4th day of October, 1940.

[Seal.] OLAVA H. GULLORD,

(My commission expires Oct. 24, 1943.)